Study of Applying Geo-spatial Metadata to Digital Map Quality Control

Chuncheng Yang

(Xi'an Research Institute of Surveying and Mapping, Xi'an, 710054)

ABSTRACT The quality of digital map is the most concerned issue for geo-spatial data users. Geo-spatial metadata are "data about geo-spatial data". The quality of geo-spatial data is one of the key comprising parts in metadata. Through our extending quality describing information of geo-spatial metadata, adding quality controlling information, and combining quality describing information with data discovering information, a geo-spatial metadata database and a digital map database have been constructed based on Microsoft's OLE DB technology firstly, in which non-spatial data and spatial data are managed integrally. Secondly, a schema of digital map quality visual inspection has been designed according to characteristics of digital map. Finally, a demo system used for digital map quality inspection and control under LAN has been implemented based on Microsoft's COM technology and client/server technology.

1 Introduction

The quality of digital map is the most concerned issue for geo-spatial data users. In the process of digital map producing, the error between geo-spatial data and real-world entity exists inevitably because of digitizing error or other factors. If the error is so large as to give false analytical results, it will give negative effect to the users' decision making, data producer's figures and production's authority. It is necessary to control digital map's quality in digital map producing when constructing map database or geographic information system.

Geo-spatial metadata are "data about geo-spatial data". They describe the content, quality, applied condition, and other characteristics of geo-spatial data. It can be used in different aspects of geographical information systems, for example, the ability to create and use a lineage meta-database model of the spatial database opens up the possibility of including a more powerful time/space tradeoff function to select an optimal configuration of derived layers to leave in the spatial database (Lanter 1993;Lanter 1994);The formalized knowledge of geographical analysis methods could also be made available in a meta-information system or knowledge base so that users could be advised directly of the best way to achieve their aims （Burrough,1992）.From the functional view, metadata can be viewed as a resource serving four functions: in data discovery; in data transfer; in data management; and in data use, although, currently, the use of metadata to support the latter two functions is less well recognized (David Medyckyj-Scott，1996). In fact，it can play a key role in data management, practically in quality management of the process of digital map producing of metadata.

For users, knowing none or a little information about the quality of data set, they wouldn't decide whether or not the data could be used in their project. Data collectors could greatly assist users by providing information about the quality of the data collected so that rational decisions can be made about whether or not include any particular data in analysis（Burrough 1992）.The people who deposit the data sets enter the meta-information at the same time as they provide the data set to be stored（Ruggles 1991）. Nowadays, digital map is produced under LAN. In order to coordinate different working procedures and control semi-finished products' quality effectively, it is a feasible and effective method to construct geo-spatial metadata database and extend it's quality description information. At the same time，it comprise the basic quality information which is useful to formalize the final quality information of a data set which can be queried by users.

Through our extending quality describing information of geo-spatial metadata, adding quality controlling information, and combining quality describing information with data discovering information, a method of constructing geo-spatial metadata database and digital map database, in which non-spatial data and spatial data are managed integrally, based on Microsoft's OLE DB technology and client/server technology under LAN has been discussed in this paper. The technology of using geo-metadata to trace, inspect and control the quality of digital map in the process of producing or final product has been introduced too.

2 Producing model of digital map and data flow chart

Digital map is a kind of high technological product. The process of producing includes the design of technological plan, the preparation and collection of source materials, the transformation of data format, vectoring, checking up data quality and other working procedures. There are quality control problems in every procedure, for example, in the stage of designing the technological plan selecting what kind of projection method, conforming to which specification should be decided；in the section of vectoring the position accuracy should be restrained in how much will not affect the quality of final product. Figure 1 shows the producing model of digital map and figure 2 shows the data flow chart between different procedures and management departures.

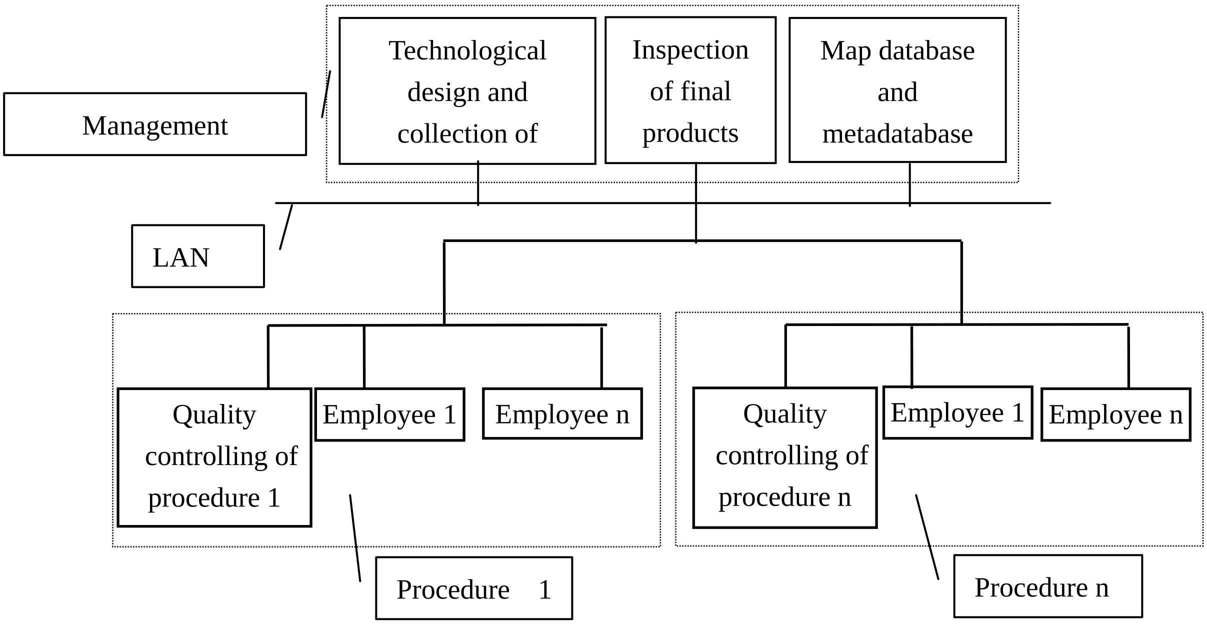

Figure 1. The producing model of digital map.

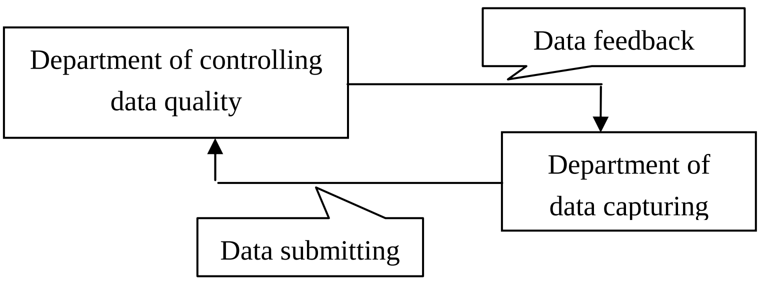

Figure 2. The data flow chart between different procedures and management departures.

3 The design of data model of map database and metadatabase

3.1 The design of data model of map database in which non-spatial data and spatial data are managed integrally

Traditionally, the spatial and non-spatial data are managed respectively：non-spatial data(attribute data) are managed by relational database such as Oracle or SQL Server and spatial data are managed in the form of file system. It involves unique identifiers stored in a database table of non-spatial that allows them to be tied to individual spatial elements. Adopting this method is an alternative in the circumstances because that the support to vary-length data is limited in RDBMS. The general performance was affected enormously because of the use of file system and frequent connection between spatial and non-spatial data. With the development of database technology, the support to vary-length data is enhanced in different database systems. It is possible to manage spatial data and non-spatial data integrally.

Nowadays, the organization of spatial data can be divided into two categories: in the first category the area entity is produced indirectly through topological relationships between linear entities and in the second classification the coordinates which construct the area entity is stored directly, such as the Arc/Info shape file format data and in MapInfo export format data. The former can reduce the space stored by data. But it is time-consuming in organization of spatial data. Contrary, the latter can improve the organization of spatial data. But the data redundancy is exist. Considering the different requirements of users and various applications in distinct discipline，the system designs a kind of data structure which can combine these two category organizations into one. The method of organizing data is showed in figure 3.

Identification code	descriptive data	topological data	visible parameters	geometry data

Geometry range of entity	multi-media information	

Figure 3.Organization of map data

In figure 3，identification code is an integer which is used to stand for an entity，equally to the primary key in database，and it is sole in a single feature coverage of a sheet of map. Descriptive information includes the descriptive code defined what the feature is (for example, highway) and parameter information (for example, the bridge is 6 meter width). Topological data include the topological information in relation with point entity, information in relation with linear entity and information with area entity. Visible parameters comprise of the scale range , perimeter , area and other geometry factors of a feature which can be used to justify if the feature is visible or not. Geometry data mean that the coordinates information which make up of a geometry point, line or area. Geometry range of entity is the minimum rectangle boundary of an entity, which plays a active role in speeding the display of electric map and queries of digital information. Multi-media information is images, labels, texts, audio and video information which is related with spatial entity.

3.2 The design of metadata model

The distinction between data and metadata is quite complex. Some author (Johnson 1991)divided into six layers：the real world、sampling on which map sheet is based (database)、content of the map sheet、map sheet description、map series、map index. He concludes that the capture of metadata is easier and more readily achieved at the higher layers and can capture of different metadata at different levels; the links between records entered at different levels are built automatically.

In this system, the organization of metadatabase includes three level: they are data set level, sheet map level and layer level respectively. In the level of data set the metadata that includes the general quality of data set, lineage information, spatial reference information, entity and attribute information and dispatching information all are used to describe the general situation. In the level of sheet map the metadta comprise of the source material, the projection methods, the standard and specifications used in producing digital map, employees who participate the project, work man, work day, the statistics information of the quality of this sheet map etc. In the level of layer the metadata include the attribute accuracy information, position accuracy, logical consistency report, and completeness report information, Figure 4 shows its constitutes:

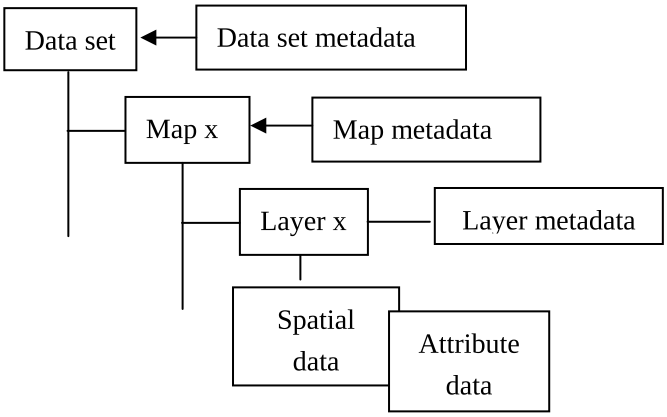

Figure 4

The metadata can be recorded in the form of documents or stored in database. Therefore, in this system the function of metadata mainly drop in quality control of data producing. It will be queried at random and be queried in different ways. In order to meet these requirements, the metadata stored in database should be convenient and flexible. The tables of metadatabase include data set metadata tables, sheet map metadata tables and map layer metadata tables, which are showed in table 1， table 2 and table 3 respectively.

Data set 1	The general quality	Inheriting information	Spatial reference information	The table names of sheet maps
Data set 2				

Table 1. Data set metadata table

Sheet map 1	Resource materials	specifications	Employees and dates	The table names of map layers
Sheet map 2				

Table 2. Sheet map metadata table

Record 1	Is this entity in map database ?	Attribute accuracy	Position accuracy	Logical consistency	completeness
Record 2					

Table 3. Map layer metadata table

In all these three levels of metadata, the data set metadata is used to describe data set macroscopically in order that users get information of data set easily. The sheet map metadata and map layer metadata are used for quality controlling and data management of digital map producing. In the layer metadata table ,the item record is used as a connection between metadatabase and map database. Once a error occurred in a procedure with an entity in digital map producing a record will be recorded in map layer metadata table accordingly. If an error of completeness occurred it will be recorded in metadata table too. The difference between this kind of error and other kinds of errors lies in the second item in map layer metadata table should give a value true or false according to different situation.

4 The design and implementation of map database and metadatabase based on OLE DB technology

4.1 The definition of digital map quality metadata item

The quality metadata of digital map includes four parts which are position accuracy, attribute accuracy, logical consistency and completeness（FGDC,1994;Zhao 1998）, in every part there are many quality control metadata items.

Position accuracy is a main factor to assess the quality of geo-spatial geometry data. It is defined by how well the true measurement of an entity on the earth's surface match the same object stored as a series of digital coordinates in a GIS data layer. The position metadata includes both horizontal position and vertical position accuracy items.

Attribute accuracy refers to how well the attribute of digital map is met with classification specification of entities. The attribute of point entities, line entities or area entities is right or not. It can be further divided into entity identification, quality description and numeric description. The entity identification is used to differentiate this entity from other entity. The quality and numeric description is used to describe the entity in details. errors occurs with these three kinds of data, so the attribute metadata includes three kinds of metadata items.

Logical consistency refers to the relationships between entities is right or not. It comprises of the consistency of topological relationships, geometric data and labels. The consistency of topological consistency includes the relationships of point entities to point entities, point entities to line entities, point entities to area entities, line entities to area entities is right or not. The consistency of geometry data includes the following situations: is the intersection between two line entity at the right place or not? Do repeat line entities exist? Does every polygon possess a interior identification point? Is the line entity too short or too long to meet another line entity? Does too little area entity exist and so on. Is the consistency of label refers to the label right in itself? Is the position of label right or not? and are the relationships between label and geometric data right or not etc. All errors fall into above situations will be recorded in metadatabase as logical consistency quality metadata

Completeness includes the spatial completeness and classification code completeness.

We defined these metadata as extending quality controlling metadata. It will be stored in metadatabase in the form of documents, number or alphabetic symbol （FGDC,1994;Shen 1999）.

4.2 The visual quality inspection of digital map and the capture of metadata automatically

The quality inspection of digital map is an important but boring work. How to find the quality problem of digital map efficiently and capture the extending quality metadata items automatically is the key problem to digital map producing. Visual technology provides us with a flexible solution to this problem.

In order to check the position accuracy, an image of sheet map can be used as a background. Through comparing with the image the error of position can be find easily and efficiently. Once inspecting the attribute accuracy, on the one hand, different graphic symbol can be used to stand for different attribute data. In this way the problems of quality of attribute accuracy can be checked out. On the other hand, letting single map layer or two map layers of digital map displays on the screen is an efficient way too.

Verification of logical consistent such as topological relationships among entities, mainly inspection of relationships between point entities and line entities and relationships between line entities and area entities, includes batch inspection and interaction inspection of human-machine two ways. Once verification in batch way, the system puts different color circle symbols on the node according to this node connecting with how much arcs and fill the area entity if its topological relationship is right. If checking up in interaction way, you can click the entity you are interested in and the topological relationships of this entity will show on the screen.

Another example is the inspection of general quality of contours, the system constructs a three dimension electronic map automatically and overlay the contours on it. Through comparing the errors of contours can be find efficiently.

The visual verification method can help finding the quality problems of digital map and stored related metadata into metadatabase automatically. The employees who are responsible for data capturing can query the metadata quickly and rectify the problems in time . In this way, the quality of digital map are guaranteed.

4.3 Implementation of map database and metadatabase based on OLE DB

OLE DB is a database scheme based on COM (Component Object Model) technology. It extends the functions of ODBC in two aspects：firstly，OLE DB provides an interface for database programming (in fact, it is COM)；secondly，OLE DB provides an interface which can connect relational or non-relational data sources.

On the book of OpenGIS Simple Features Specification For OLE/COM ver1.1 published by OGC(Open GIS Consortium), it said the OLE DB is an ideal exposing interface for geo-spatial data handling, for the integration of geo-spatial data and non-spatial data is easy and flexible (Open GIS Consortium).

Comparing with ODBC the advantages of OLE DB lie in following aspects(Robison,1999)：firstly, OLE DB use COM interface to connect with data source，while ODBC through traditional DLL and import library's static links； secondly, OLE DB separates the setting of user identification, password and DSN with the data source，So it is easy for application to persist its connecting attributes；thirdly, OLE DB is a complete object-oriented method， while ODBC employs application programming interface.

Combining OLE DB with ODBC technology as a database client development tools and employing SQL server as a database management system, this system resolves the problem of management of spatial data and non-spatial data integrally through the help of supporting to vary-length data of OLE DB and SQL server. The details of implementation shows in figure 5.

```
//STEP 1----Data binding
//ODBC type
SQLBindParameter(hstmt,1,SQL_PARAM_INPUT,SQL_C_BINARY,SQL_LO
        NGVARBINARY,MAX_XY_LEN,0,XY,MAX_XY_LEN,lvXYInd);
//OLE DB type
DBBINDING     dbBinding[2];
dbBinding[0].iOrdinal        = 1;
dbBinding[0].pTypeInfo       = NULL;
dbBinding[0].pObject         = NULL;
dbBinding[0].pBindExt        = NULL;
dbBinding[0].dwMemOwner= BMEMOWNER_CLIENTOWNED;
dbBinding[0].eParamIO        = DBPARAMIO_INPUT;
dbBinding[0].dwFlags         = 0;
dbBinding[0].bPrecision      = 0;
dbBinding[0].bScale          = 0;
dbBinding[0].cbMaxLen        = MAX_XY_LEN;
dbBinding[0].wType           = DBTYPE_BYTES;
//STEP 2----Insert into database or Query information from database
………
```

Figure 5. The details of implementation of map database and metadatabase

5 Conclusions

This paper illustrates that through our extending quality describing information of geo-spatial metadata, adding quality controlling information, a geo-spatial metadata database has been constructed based on Microsoft's OLE DB technology. Linking with map database, it can control the quality of producing procedure of digital map efficiently. A visual verification and controlling subsystem has been constructed under LAN in order to inspect the quality of digital map and capture the quality metadata automatically. The metadatabase can help data providers to improve its quality of digital products.

References

B D Hohnson etc,1991, The FINDAR Directory System: a meta-model for metadata. Metadata in the Geosciences, Group D publications ltd.

Cliver Ruggles and Ian Newman, 1991,The MRRL's Meta-Information Retrive and Access System. Metadata in the Geosciences, Group D publications ltd.

David Medyckyj-Scott etc, 1996, Discovering environmental data: metadatabase, network information resource tools and the GENIE system. INT. J. Geographical Information Systems, VOL.10,No.1,65-84.

David P. Lanter, 1993, A Lineage Meta-database Approach Toward Spatial Analytic Database Optimization. Cartography and Geographic Information Systems,Vol.20,No.2, pp112-121.

David P. Lanter,1994, A Lineage Metadata Approach to Removing Redundancy And Propagating Updates in a GIS Database, Cartography and Geographic Information Systems,Vol.21,No.2, pp91-98.

Federal Geographic DataCommittee (FGDC 1994,rev 1997).Content Standards for Digital Geo-spatial Metadat. http://www/fgdc.gov/metadata.

Lyn Robison,1999,SAMS Teach Yourself Database Programming with Visual C++ in 21 Days. Publishing House of Electronics Industry, Beijing.

Open GIS Consortium, Inc. OpenGIS Simple Features Specification For OLE/COM.Revision 1.1.

P.A.BURROUGH,1992, Development of intelligent geographical information systems. INT . J. Geographical Information Systems,VOL.6,No.1,P1-11.

Shen Tiyan and Cheng Chengqi,1994, Design and Realization of Spatial Metadata System. Journal of Wuhan Technical University of Surveying and Mapping,No.4.

Zhao Yongping,1998,The Research of Metadata Standard Based on NSII and The Establishment of Experimental Sharing System. Dissertation of Peking University.