A Construction Schema for Provincial Spatial Database of China

WANG Yandong  GONG Jianya

Abstract  In order to provide a provincial spatial database, this paper presents a scheme for spatial database construction to meet the needs of China. The objective and overall technical route of spatial database construction are described. The logical and physical database models are designed. Key issues are addressed, such as integration of multi-scale heterogeneous spatial databases, spatial data version management based on metadata and integrative management of map cartography and spatial database.

Keywords  spatial data; database; construction; provincial

CLC number  P208

Introduction

At present, the fundamental geographic information data of 1:10 000 scales are produced, managed and distributed by provincial bureaus of surveying and mapping in China\[1\]. Each provincial bureau of surveying and mapping has constructed or is constructing its own provincial spatial database. However, there is no integrated schema for the construction of a provincial spatial database. This paper puts forward a scheme which describes the whole process of provincial spatial database construction, including the objective and overall technical route, the logical and physical database models and key issues.

1) The main objectives of spatial database construction are illustrated follows.

In China, the spatial data above 1:1 000 000 scales are classified and the distribution of these data are subject to various restrictions. Therefore, the provincial spatial database can only be constructed over a network that is physically isolated from the Internet. The main objective is for management and distribution.

- Build spatial database management systems and achieve integration of spatial data from different sources.
- Build a sharing mechanism and a distribution service system of fundamental geographic information data.
- Build a good system of data service, shorten the traditional way of “data retrieve—data order—data process—data offer”, and enable users to “find—see—take” the data.

2) Analysis of business flow of spatial data management and distribution

A fundamental geographic information database is the “total window” that provides the fundamental...
geographic information services to the community and the “main node” by which the fundamental geographic information enters the national information highway network[2]. Fig.1 is the business flow chart of spatial database management and distribution.

Users can query the information they need from the metadata-database released on the Internet and then send the trade request. When a trade request is received, a data distribution clerk retrieves the corresponding data in the fundamental geographic information database, confirms the required data and then submits the request to a data manager for the data requisition. According to the request, the data manager sends the corresponding data to the data distribution clerk. Finally, the data distribution clerk sends the data to the user after he verifies the data and transacts procedures with the user.

3) Overall technical route of database construction

After defining the overall goal of building a database, the logical and physical model of the database are designed according to the analysis of user’s demand of the fundamental geographic information database and the spatial data itself. Data storage is built according to the design scheme. Spatial data are input into the database after being checked and processed. When databases are integrated as one, the whole system is tested. After being checked and accepted, the construction of the database is completed. Finally, the work turns into daily running and maintenance of the database. Processing details are shown in Fig.2.

![Fig.2 Flow chart of database construction](image)

### 1 Database design

#### 1.1 A logic model design

Different data models are adopted according to different types of data in the construction of the provincial fundamental geographical information database[3].

- The vector database is modeled as object-oriented or object-relation model.
- The DOM database and the DEM database are modeled as pyramid model. The vector database, DOM database and DEM database are beyond the concept of the map sheet and built in a seamless pattern.
- The metadata database is modeled as a relational model.
- The aerial photo database and the DRG database are modeled as the relational model and catalog management.

The provincial geographic information database can
be divided into seamless database, catalog database, metadata database and cartography database as illustrated in Fig.3.

The details of logical design are shown in Fig.4.

1.1.1 Seamless database

It is built for the sake of on-line searching, browsing and analyzing. Many technical means are used to process data, such as data conversion, projection transformation. It is a logical or physical seamless database based on the spatial database management system or geographic information system. It provides efficient search, distribution and analysis service for users. This connects various scales and various types of spatial data together, and an integrative spatial database is formed. There are temporal database and historical database in the seamless database. The temporal database saves the updated data. If there are two or more versions of spatial data in the same data unit, the pre-version of spatial data becomes historical data after the data are updated. The historical data is an important data source for inspecting the change of nature.

1.1.2 Catalog database

Because of the irregularity of the data area and high overlap, the aerial photo and DRG are not suitable for seamless organization. However, they are suitable for management by catalogs and being stored in the database in the form of single data file according to a certain geographic area. The aerial photo and DRG managed by catalog can do overlay with the data in the seamless database. Furthermore, they can be located and searched with the place-name database and vector database and then be distributed.

1.1.3 Meta-database

The descriptive information of fundamental geographic data, including the information of data production and data management, are stored in meta-database. Meta-databases of product level, database level and map sheet level need to be established. The meta-databases can satisfy various query and search requirements in data production, management, and distribution.

1.1.4 Cartography database

To resolve the inter-zone problem, vector database is established by geographic coordinate systems. However, the spatial data based on the geographical coordinates system is not suitable for outputting China standard topographic maps. Also, the data must be symbolized and processed according to the production standards before outputting standard topographic maps. The cartography production database is established to avoid the modification of fundamental geographic information database. Data are extracted from
fundamental geographic information database to production database, and the projection transformation of the data is done. A cartography production database is in the form of file or ACCESS database and managed by map sheet. The data are edited, symbolized, groomed, printed and outputted in the cartography production database.

1.2 Physical model design

Attribute data should be of a uniform data structure. Features of a certain type correspond to an attribute table, and an attribute table can include many types of features. Attribute data structure design includes identifying the names, types, fields and attribute values of the attribute items of certain features. Table 1 gives attribute item names and definition of 1:10 000 scale DLG data.

| Field | Input width | Output width | Data type | Decimal digits | Annotation |
|-------|-------------|--------------|-----------|----------------|------------|
| CODE  | 3           | 3            | I         | -              | Temporary code |
| GB    | 5           | 5            | I         | -              | International code |
| ELEV  | 8           | 8            | N         | 2              | Elevation code |

The attribute table content and order of attribute items in each layer are in Table 2.

| Layer name         | Attribute name                          |
|--------------------|-----------------------------------------|
| CTRPTP(control point) | CODE, GB, NAME, LEVEL, ELEV, XCOOR, YCOOR |
| HYDLKL(water system) | CODE, GB, TN, NAME                      |
| HYDLKP(water system) | CODE, GB, TN, NAME                      |

2 Database construction

2.1 Construction process

Database construction includes database building, data checking before input, data processing, data input, data checking after input, etc. The database construction process is shown in Fig.5.

2.2 Database storage construction

The database storage construction includes physical space allocation, parameter setting and data table creation, etc. according to the physical design and logical design of the database.

2.3 Data checking

In order to ensure data quality, before data input, spatial data should be checked according to production technology and quality specifications. If it cannot meet the requirements, the spatial data should be rectified. The data to be checked includes DLG, DOM, DEM, DRG and metadata. The checking contents of DLG, DOM, DEM and DRG can refer to the first degree quality elements in PRC specification GB/T 18316-2001. The checking of metadata and place-name data includes data format, data item integrality, data item naming and data content normalization. The checking requirement of each type of data is listed in Table 3.

2.4 Data processing

There are two types of spatial data processing. The first one is the processing of the raw data or the data
that does not meet the quality requirement. The other is the processing of the data stored in the database from which other useful information can be derived.

Provincial fundamental geographic information database is mainly responsible for spatial data management and distribution. Therefore, the data processing must satisfy the following requirements.

- The mathematical accuracy of data cannot be lowered.
- No data is lost.

| Type       | First degree elements                          | Second degree elements                          | Data to be checked                                                                 |
|------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------------------------------------------|
| DLG        | Basic requirements                            | Property accuracy                              | The data formats are Arc/Info E00, DWG/DXF, Mapinfo (Mif/Mid), GeoStar GWS and VCT etc. |
|            | File name, Data format, File number            | Property structure                             | Metadata format is mdb database.                                                  |
|            | Basic requirements                            | Metadata                                       |                                                                                    |
|            | File name, Data format, File number            | Map scope of specified coordinates and tolerance |                                                                                    |
|            | Basic requirements                            | Basic requirements                            |                                                                                    |
|            | File name, Data format, File number            | Math foundation                               | DEM data format includes ArcInfo bil, GeoStar dem, national specification exchange format grd etc. |
|            | Mathematic accuracy                            | Grid spacing                                  | Metadata format is Access database.                                              |
|            | Classification correctness and integrity of feature layer | Expanding scope                              |                                                                                    |
|            | Metadata                                       | Correctness and integrity of metadata          |                                                                                    |
|            | Basic requirements                            | Basic requirements                            |                                                                                    |
|            | File name, Data format, File number            | Math foundation                               | DOM includes image data files (tif, bmp, Geotif etc).                             |
|            | Mathematic map accuracy                        | Resolution, Expanding scope                   | Metadata format is Access database.                                              |
|            | Metadata                                       | Correctness and integrity of metadata          |                                                                                    |
|            | Basic requirements                            | Basic requirements                            |                                                                                    |
|            | File name, Data format, File number            | Math foundation                               | DRG includes image data files (tif, bmp, Geotif etc).                             |
|            | Mathematic accuracy                            | Math foundation                               | Metadata format is Access database.                                              |
|            | Metadata                                       | Correctness and integrity of metadata          |                                                                                    |

- The integrity and consistency of vector data cannot decrease.
- Image data quality does not decrease.

Spatial data processing includes many contents. Among them the most important two contents are spatial data conversion and adjustment between map sheets. Spatial data conversion includes code conversion, data format conversion, coordinates conversion and projection conversion. Spatial data adjustment includes the following contents.

- Vector data adjustment between map sheets, which includes feature property and graphic adjustment.
- Image adjustment is mainly responsible for hue adjustment between map sheets.
- DEM adjustment is mainly responsible for elevation adjustment between map sheets.
- After data adjustment between map sheets, databases should be logically seamless, the correlation between the databases is true and the feature property is consistent.

2.5 Data input

Spatial data input can be divided into input by region and input by feature. The former manages and stores data by region or map sheet, the latter manages and stores data by feature layer. Input by region pertains to image data and DEM while input by feature pertains to vector data.

2.6 Database system integration

In database system integration, data integration, hardware and software integration, and function development and integration must be taken into consideration. Data integration refers to the establishment of relations between data of different types, data with the
same scale, data with different scale, and metadata and corresponding data. Only by the establishment of these relations can data management and distribution be carried out smoothly. Data integration includes the following contents.

- Integration of different types of spatial data. Vector, raster, DEM and some other kinds of data can be flexibly integrated with each other, for instance, overlay display of different types of spatial data, and spatial analysis, on the data integration platform.

- Integration of data with the same scale. That all data with the same scale are stored in the same table in the database ensures the data to be integrated seamlessly.

- Integration of data with the different scale. For instance, raster data with different scale can be integrated flexibly by pyramid.

- Integration of metadata and their corresponding data. Achieve a two-way query between metadata and their corresponding data by building the relationship between them.

3 Key technologies

The key technologies involved in database construction include integration of multi-scales heterogeneous spatial databases, inter-zone panning of raster data, spatial data version management based on metadata, integrative management of cartography and spatial database and management of catalog data.

3.1 Integration of multi-scale heterogeneous spatial databases

Because of different data forms among these data sources, we usually choose different methods for different data types to construct the database. Although spatial data are numerical and can be displayed in multi-scales by zooming in or zooming out, it is produced according to survey specifications of different scales. The data of each scale have their own data precision. The data are accepted or rejected in different scales. It is called resolution.

Because the automatic synthesis of map data has not been solved completely, at present, it is impossible to build spatial databases of a certain scale from which spatial data of random scale can be abstracted. Therefore, we need to produce spatial data of different scales and build a multi-scale spatial database separately. When we use the object-relation database management system to manage vector data, the spatial data of different scales can be stored in the same database, but must be separated in logic. In addition, in order to establish the relations among the data of different scales, a unified displaying and dispatching mechanism need to be established. When we build spatial data databases, the data of different scales may have different spatial references. For example, a coordinate system of 1:500 00 data adopts 6° Gauss coordinate system and 1: 100 00 data adopts 3° Gauss coordinate system. The data are transformed to the unified coordinate system by fast coordinate transformation when displaying.

3.2 Spatial data version management based on metadata

Metadata are the data about data. Metadata can help organize, manage, search and distribute spatial data rapidly. In the work of the provincial spatial database, a data version management scheme based on metadata is put forward. Each map sheet includes spatial data and their metadata in which the data contents, quality, state and other characteristics are described as text file or other format.

The spatial database can be divided into three sub-databases logically: current spatial database, historical spatial database and metadata database. In the current database, the most updated spatial data are stored. In historic database, various version data are stored, and metadata of all versions of map sheets are stored in the metadata database. Based on metadata, the input, updating, scheduling, searching and distribution of various versions of spatial data are implemented.

3.3 Integrative management of cartography and spatial database

The provincial geographic information database system is able to produce digital maps. It can do integrative management of cartography and spatial database, producing standard topographic maps and various thematic digit maps by using the spatial data in the database directly.
To solve the problem of data cross zones, the geographical coordinate system is used for the construction of the fundamental geographical information database. However, the geographical coordinate system is not suitable for outputting Chinese standard topographic maps. In order to gain a beautiful effect, the data must be symbolized and edited, such as overlay adjustment before output. To avoid destroying the original data, the cartography database is built. In the cartography database, the data need to be extracted from the fundamental geographical information database to production database, and data projection is required in this process. The cartography database is in the form of ACCESS database in which data is stored by map sheet. Fig.6 shows the flow chart of the cartography database.

### 3.4 Management of catalog data

Some kinds of data in the provincial fundamental geographical information database, such as aerial photos and DRG, etc., are not suitable to be organized by a seamless form in the database. For example, the aerial photos have a big overlay between adjoining photos because the airplanes fly by flight strip. Besides, the aerial photos are not rectified and enchased. Therefore, the aerial photos cannot be managed by the rules such as the standard map sheet. It can only be managed by a catalogue form.

- The aerial photos are managed in the unit of the photography area. Each photography area sets a work area and has a series of instruction documents; each aerial photo has a photo control point which notes the central geographical coordinates of the aerial photo. The extension of each aerial photo can be worked out from the control point, the resolution and the pixels of it. When the aerial photos are put into the database, each photo is a unit. At the same time, the extension and the storage in the database of the photo are saved in the catalogue. The instruction document of each photography area is saved into the database as the metadata.

- The database system offers the functions of overlay aerial photos with other data in the database, locating a place by its name, querying the number of aerial photo by its extension, displaying a specified aerial photo and extracting the aerial photos etc.

### 4 A case

The spatial database construction scheme introduced above has been adopted in many provinces in China (Shanxi, Fujian, Hubei, etc.). It can meet the requirement of Chinese province fundamental geographic information database construction. We take the database construction of Shanxi province as an example.

![Fig.6 Flow chart of cartography database](image)

GeoStar 5.0 which is a GIS software in China is selected as the GIS platform and Oracle 9i is selected as the database platform. The data size of Shanxi province is approximate 700 GB.

Fig.7 is the overview of Shanxi province. In the user interface, data that are different in scale, accuracy and size can be overlaid to display seamlessly. Fig.8 shows a map of fundamental scale with Chinese characteristic. The provincial bureau of surveying and mapping develops the specifications for symbol and map output. In the specifications, the symbol style of the map with various scales is specified. The specifications include every detail of map decoration (map name, index to adjoining sheets, legends, etc.). Fig.9 is DLG data

![Fig.7 The overview of Shanxi](image)
Spatial data can assist in decision-making in many areas, and more and more industries begin to use spatial databases. In the vigorous developing context of China, the application of spatial data is at an initial stage, but the requirements are increasing intensively because of its advantage and the development of infrastructure. Owing to the characteristics (variety, multi-scale and large volume, etc.) of fundamental geographic information, a spatial data management and distribution system must be developed to meet the needs of China.

5 Conclusion

Spatial data can assist in decision-making in many areas, and more and more industries begin to use spatial databases. In the vigorous developing context of China, the application of spatial data is at an initial stage, but the requirements are increasing intensively because of its advantage and the development of infrastructure. Owing to the characteristics (variety, multi-scale and large volume, etc.) of fundamental geographic information, a spatial data management and distribution system must be developed to meet the needs of China.

References

[1] Gong Jianya, Du Daosheng, Li Qingquan, (2004) Contemporary geographic information technologies[M]. Beijing: Science Press

[2] Shanxi Bureau of Surveying and Mapping (2006) Overall design of fundamental geographic information database in Shanxi Province[OL]. http://other.sxinfo.net/mapdds/htm/s-zx.htm

[3] Shanxi Bureau of Surveying and Mapping (2006) Technical report of fundamental geographic information database in Shanxin Province[OL]. http://other.sxinfo.net/mapdds/htm/s-zx.htm

[4] Fang Tao (1999) The development and realization of GeoImageDB multi scales and seamless image database system[J]. Journal of Wuhan Technology University of Surveying and Mapping, 21(3):189-193

[5] Davis J R (1998) IBM’s DB2 spatial extender: managing Geo-Spatial Information within the DBMS[OL]. http://www-3.ibm.com/software/data/spatial/spatial.pdf

[6] Abel D J. SIRO-DBMS (1989) A database tool Kit for geographical Information systems[J]. Int. J. of Geographical Information Systems 3(2): 103-116

[7] Günther O (1988) Efficient structures for geometric data management[J]. Lecture Notes in Computer Science, 337, Springer

[8] Oracle Corporation (2006). Oracle spatial: an oracle technical white paper[OL]. http://www.oracle.com/technology/documentation/index.html.