Integration and Sharing of Geo-spatial Data Based on Data Engine

WU Xiaofang, CAI Zhongliang, WU Guofeng, DU Qingyun

1 Introduction

At present, the source of GIS spatial information is very extensive. The data to be processed with most popular GIS software are Arc/Info data, MapInfo data, GeoStar data, MapGIS data, CorelDraw data and so on.

The vector transfer format (VCT) among geo-spatial data transfer format is a normal one to implement the sharing of data\(^1\). It can realize the sheer transfer of all kinds of data and ensure the integrity of information.

Nowadays the integrating modes of present mapping data are mainly the transfer of data formats, but its shortage is that the users can not share all kinds of GIS data seamlessly. Therefore, this paper presents the method for accomplishing the integration and sharing of all kinds of GIS vector data by using data engine method on the basis of geo-spatial data transfer format.

2 Data integration

The difference of various GIS data formats takes much trouble in integrating and sharing information. On the other hand, it also provokes the software developers to try to resolve the problem of data sharing. Presently, for implementing the integration and sharing of multi-format data there are three methods\(^2,3\) : the transfer mode of data formats, the interoperable mode of data and the mode of directly accessing the data.

2.1 Transfer model of data formats

At present, the majority of GIS software in China adopts the transfer mode of data formats. Users read and write...
the inner format, external format or some normal format of GIS software on the basis of understanding the specification of its format, then transfer it to the acceptable data mode and realize data sharing.

Although this method can resolve the problem of data sharing to a certain extent, it is difficult to really fulfill the data sharing owing to the difference in representation and emphases of all kinds of data formats. And it is easy to lose the information during exchanging with each other. At the same time, if the version of the data format is upgraded, the transfer module must be rewritten. Therefore, the method is an inferior sharing method, not good in opening.

2.2 Interoperable model of data

The interoperable mode of data is based on the OpenGIS standard. It makes GIS users access their necessary information transparently after understanding each other. The client/servers of different platforms realize data sharing by uniform protocol.

The interoperable mode of data brings new methods for integrating data, but there is certain limitation in practical application. First of all, in order to accomplish the interoperability between all kinds of data indeed, GIS software which provide data formats must keep the uniform specification to fulfill data accessing interface. However, it can not be achieved sometimes. Besides, one software has accesses to the data formats of other softwares by means of the data server. This server is the host software of data format accessing. However, it can not be achieved sometimes. Besides, one software has accesses to the data formats of other softwares by means of the data server. This server is the host software of data format accessing. Lastly, at present the majority of GIS data is the spatial data formats of non-OpenGIS standard. Even if new GIS softwares support the OpenGIS standard, the spatial data produced by current GIS softwares still need transferring to OpenGIS standard.

2.3 Mode of direct data access

The mode of direct data access is to realize data format access of other software in one GIS software. User can use one GIS software to store and read several of data formats. The mode of direct data access not only avoids fussy data transfer, but also need not have the host software and execute the software when one GIS software accesses the data format of host software.

The mode of direct data access must be based on the full understanding of accessing data formats. If the accessed data format is not open, this format must be unscrambled firstly. At the same time, we must ensure that this format is unscrambled accurately. Thus the complete sharing of all kinds of data can be accomplished. If the formats change, we have to unscramble new formats by providing new upgraded version. On the other hand, when the formats change, they are usually not announced immediately. Therefore, other data software which are based on these GIS data formats will fall behind.

3 Data integration based on data engine method

Reviewing the principle of the share of multi-source data in database systems, the authors find out that ODBC (open database connectivity) resolves the incompatible problem within the databases by adopting a normal method to access certain database. The database software only need add the ODBC interface to their database language and developing tools and make the database open.

Therefore, the authors present that using data engine method, just as the driver in ODBC provides dynamic link libraries, we can set up the logical data model of the data source of all companies on the basis of VCT. Only if the drivers support this logic model, in some applications SQL and API function can be used to send over the access request, such as managing the existent physical data model. With the data engine method, the drivers dynamically transfer the received request to the command format which the data source can
accept, then extract the necessary data information (Fig. 1).

![Diagram](image URL)

**Fig. 1** Logical structure of the data engine method

### 3.1 System architecture

Constructing the data engine is the key to realize data sharing. The system architecture is described as follows:

1. Define an abstract data model which includes all features of spatial data (VCT can be fit basically), then define an abstract base class for storing this model in OOP language, which must include the pure virtual functions (the public functions are open to all kinds of GIS software), such as reading basic objects, operating basic objects, etc.

2. For the data of every GIS software, derive an inherited class of certain special format from the base class in order to implement operating the special format, and store it by means of dynamic link libraries. Every GIS software only need finish its own derivative classes, but not change its format or produce the redundance of data. At the same time, when the data formats of GIS softwares change, it is OK to provide the latest dynamic link libraries and it does not affect the sharing of other data. It is must be pointed out that when a certain format has not the function that the abstract data model defines, the function of this derivative class must return default value.

3. It is necessary to provide the meta file including the defined formats of all functions in the derivative classes of dynamic link libraries, the specification of parameters, the files which explain the result after running.

### 3.2 Representation of basic features in VCT

In VCT, two-dimensional features have five types: point, line, surface, text and complex object. Their representations in VC++ are described as follows, respectively.

1. Point indicates the geometric center or the center of gravity of the symbol, and can be represented by coordinates \((x, y, z)\).

2. Line represents the features such as roads, rivers, etc. Line is made up of the continuous coordinates and is noted as \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\).

Because the types of lines are different, all their representations are not completely the same. If we transfer the curves to segment lines by means of an approximated representation, the precision will be lost. Therefore, as for the special curves, we can use different representations to store the characteristic parameters. At the same time, considering some curves which are made up of different special curves we can firstly store the special curves which made up the group object, and record their IDs, then store all their IDs.

3. Surface is made up of the continuous coordinates. According to the difference of line types, we can classify them as ellipse, pie, round rectangle, bezier curve, polygon and so on. As to the different types of surface, we adopt the different storages, and its representation is similar to the lines' representation. So we can reference the representation of line.

4. Text represents the names of features such as rivers and mountains. It involves in many attributes such as size, style, color, italic, bold, etc.

5. Complex object is made up of some simple objects. We can use the method mentioned above to organize the data. Firstly store the simple objects which made up the group object and record their IDs, then store the group object and their IDs of the simple
objects.

As for the spatial objects mentioned above, we can connect them with their attribute data by the object ID in order to implement their mutual retrieval.

3.3 Implementation of data integration

VC++ supports the software development of OOP and has many priorities such as abstraction, encapsulation, inheritance, polymorphism etc., and it meets the demand of the integration of vector mapping data.

(1) Design of the base class

Many GIS vector data have the same characters when they exchange with VCT data by data engine, so we can make full use of the advantages of OOP VC++ such as inheritance, encapsulation. When designing a base class (CTeleFile) firstly. In CTeleFile, we can define the functions such as retrieving, storing, operating and visualizing attribute data and spatial object (including the data transfer) (Fig. 2), then as far as the different data files are concerned, we can design and realize the special functions in the derivative classes.

(2) Design of derivative classes

As for every different file format, we can accomplish the advanced integration and sharing between data by realizing the pure virtue functions and exchanging data between VCT format and other formats.

As far as the usual GIS software data are concerned, drive one derivative class from the base class, then implement the functions such as reading, displaying, operating and storing mapping and attribute data in the derivative class.

Certainly, if a new GIS software is developed, we can realize the transfer between VCT and the new format only by deriving classes from the base class, CTeleFile (or providing dynamic link libraries). In this way, all kinds of the company of GIS software can bring more convenience to the developers of other software in the condition of hiding the basic source codes of some formats. Once the inner data formats change, we only need to provide the upgraded dynamic link libraries. Further more developers and the software in which the dynamic link libraries are used need not make any change or translate and edit the program. At the same time, this method can resolve the transfer between all kinds of data conveniently which can be realized in the "output file" function in derivative classes.

In VC++, the authors design the base class CTeleFile and some derivative classes in Fig. 1 by using MFC (Microsoft Foundation Class), and implement the interoperability between the usual vector data formats and VCT.

Fig. 3 is the interface for transferring the SHP file to the VCT format by using the "output file" function. Fig. 4 illustrates the visualization of AI (Adobe Illustrator) file data.

4 Conclusions

On the basis of VCT format, this paper realizes the interoperability, format transfer between usual GIS data, implements the integration and sharing of different software data.

VCT format may be in completely compatible with other formats. Therefore, besides VCT format, we may also implement the integration and sharing of data on the basis of other
Fig. 3 Interface of data transfer

Fig. 4 Visualization of AI file data

References

1 Gong J Y (1999) Some theories of present GIS. Wuhan: Press of Wuhan Technical University of Surveying and Mapping. (in Chinese)

2 Zhu X Y, Gong J Y, Huang J T, et al. (2000) The organization and management of spatial data in GeoStar. Journal of Wuhan Technical University of Surveying and Mapping, 25(2): 122-125 (in Chinese)

3 Wang Y D, Gong J Y, Huang J T, et al. (2000) The transfer method of data based on Geo-spatial data transfer format. Journal of Wuhan Technical University of Surveying and Mapping, 25(2): 143-146 (in Chinese)

4 Chen Y H (1999) The research of digital product standardization in basic geographic information. Bulletin of Surveying and Mapping. (in Chinese)

5 Chen J C (2000) Developing GIS system in VC++. Beijing: Publishing House of Electronics Industry. (in Chinese)