SOME ISSUES ON URBAN CENSUS GIS DESIGNING

WANG Xinzhou
LIU Zongwei
CHEN Shunqing

KEY WORDS urban census GIS; data warehouse; data analysis

ABSTRACT This paper first briefly looks back on the history of census GIS around the world, and then summarizes some preponderances of developing urban census GIS in China. And then a three-tier construction architecture for the urban census GIS is proposed. Finally, using the illustration of the census data of Guangzhou city, the paper analyzes and elaborates some issues on the urban census GIS designing, such as data management, data warehouse building, and data analysis.

1 Introduction

The 5th National Census of China was commenced on November 1st, 2000. It is well-known that census data are not only very necessary to serve macroscopical strategies of the state, but also extremely valuable for the decision-making of local governments and even private sectors. By using some advanced information technologies during the course of the 5th National Census, we could further take the advantage of the census information to strengthen scientific decision-making. Especially, the employment of GIS in the 5th National Census of China becomes extremely necessary and imperative in order to meet the increasing requirements of economic development for information technology.

The integration of census and geographic data can not only elevate the management of census information, but widen the application of census information in other potential fields. As a result, it will further benefit the research in demography, decision-making of population control and strategies of sustainable development in society and economy. Therefore, currently, many countries, no matter the developed or the developing ones, are attaching great importance to the study and implementation of the census GIS. Among them, the United States plays a distinguished role in the census GIS.

From 1980 to 1990, the Census Bureau of the United States developed the TIGER (Topologically Integrated Geographic Encoding and References) system, and applied it in the national census of 1990[1-3]. Integrated with statistical data, the TIGER system was widely used in many fields ranging from banking, real estate to telecommunication and postal industry. All those successful applications have turned out to be that the system has deeply influenced socioeconomy in almost every field and effectively promoted the development of information industry.

Subsequently, Japan, France, Britain, Canada and many other countries launched their own census GIS. For instance, in Britain, GIS was treated as a necessary tool for analyzing and processing census data, and in order to meet the demands of mapping and analyzing, several census GISs such as HMSO and SASPAC have been developed[3].

In recent years, China has made relatively great progress in the modernization of census information management and analysis. In the late 1990's, the Chinese Ministry of Public Security gradually begun to manage population information using com-
puter technology, and scheduled to set up the population information network system covering the whole state. However the principal functions of existing population information systems mainly focus on data statistics, and fail to precisely reflect the actual spatial distribution of population, and poorly meet the demands of the development of society and economy. Therefore, objectively speaking, the fully genuine census GIS is still unavailable in China at this time.

2 The foundations of building urban census GIS in China

2.1 Technical foundations

With the rapid development of GIS and the wide use of information techniques in China, many cities have possessed abundant techniques to develop the urban census GIS.

China has initiated the research and development of GIS as early as in the 1980s. Up to now, some practicable GISs have been accomplished. For example, the building-up of 1:1,000,000 geographic database brings about a number of technical and practical application problems on GIS. And unexpectedly, several market-oriented, excellent domestic GIS platforms emerge during this period. Roughly, the development of GIS in China can be divided into four phases which are beginning (from 1970 to 1980), preparation (from 1980 to 1985), development (from 1985 to 1995), and industrialization (from 1995 to now). The successful applications of GIS in many fields have drawn great attentions from governments at different level. And thereby, GIS as an innovative industry has been ready for industrialization.

The training and application of computer technology have been initiated by advocating and promoting non-paper official automatization (OA) in most cities since the middle 1980s. Information centers are set up to ensure the effective applications of management information systems in governmental departments such as the Urban Planning Bureau, the Public Security Bureau, etc. Therefore, the information atmosphere of E-OA and e-commerce, and relatively solid technological preparation for building up the urban census GIS have make it imperative and urgent to employ GIS technology in the efficient management and analysis of census data.

2.2 Data foundations

Taking into account the practical requirements of economy and society, the general investigation method and the sampling method were combined to use in the 5th National Census. Ten percent of families were asked to fill the “long” census questionnaires, while others filled the “short” census questionnaires. The latter contains nineteen items, while the former forty-nine items, which embody abundant information concerned with a resident’s education, age, housing condition, etc. Moreover, local and central census bureaus also inventoried a plenty of historic census data and other valuable socioeconomic materials.

In recent years, every city has also accumulated substantial achievements in the construction of spatial data infrastructure, ranging from variable scales of paper, vector and raster-based topographic maps to photogrammetric and remotely sensed satellite imagery with variable resolutions. For example, Guangzhou city possesses newly updated and comprehensive 1:10,000 vector atlas, which contain nearly forty feature layers such as road, residential house, water resource etc. Assisted by effective data-updating strategies, those data are very valuable in many spatial information-related application fields. More important, high-resolution photogrammetric and remotely sensed satellite imagery can not only extract rich urban information of the city of Guangzhou (i.e. pollution and virescence information), but also be used as an important data updating source. Partial data are illustrated in Fig. 1 to Fig. 4.

2.3 The solid social background and support of governments

In China, the development and application of information technology follows the rule of from governments to firms, and then to the whole society. Today, the application of information technology, behaving as a kind of kernel resources in the society has reached each corner of the modern society, in-
fluencing the world very widely and deeply. Especially, urban population and economy data, as the most important and essential information for a city, will not only shape the cornerstone of governmental decision-making, but also coincide with the increasing demands of the modern society.

Census is one of the greatest social mobilization in a peaceful period. This time, Chinese government backed the 5th National Census with huge financial and human resource support and naturally it draws close attention from all of Chinese people. One distinguished feature of this census is the adoption and application of a series of new technologies and methods in the procession of census data collecting and handling. Among them, the application of GIS is one of the most distinctive measures. By using GIS technology, the census data can be tightly integrated with other information and therefore broaden the application fields for the large-scale social services of census data.

With the increasingly urgent demands from the society, the accomplishment of the 5th National Census and strong supports of the government, China faces a golden opportunity of building its own Census GIS. At present, some cities, such as Beijing, Shanghai and Guangzhou, have initiated a series of census GIS-related researches and applications, and thereby have sparked a hot tide of constructing domestic census GIS.

3 Structure designing and software platform choosing

3.1 Architecture designing for urban census GIS

According to the current status of urban information level and social demands, the essential principle for the construction of urban census GIS is to exploit the scientific management and application pattern of urban census data and thereby advance the high-efficient management and application in depth by construction of the census GIS.

On the basis of the principles mentioned above, a
three-layer architecture for the census GIS designing is proposed: the basic data layer, the data management layer and the data application layer. For example, the basic data layer in the Guangzhou urban census GIS includes historical census data, urban spatial data, industry census data, and public facility data and the like from the 5th National Census. The data management layer includes the geo-population database system and population data-based data warehouse; the data application layer includes some GIS-based application sub-systems. Fig. 5 is the architecture of the Guangzhou census GIS.

![Architecture of the Guangzhou census GIS](image)

In this three-layer architecture, the data management layer is the integration of geo-demographic database system and data warehouse system. This kind of architecture is advantageous in conquering the flaws of traditional databases and can thereby manage tremendous and complex urban data (including historical data) efficiently without sacrificing the system efficiency of daily processing and database updating ability.

The data application layer includes four application sub-systems. The comprehensive application sub-system implements census and urban spatial information based on GIS functionality. The statistic transaction system provides specific statistic services regarding implementation, data processing and information services of census. On the basis of the population data warehouse, the decision support sub-system (DSS) supplies population information services by integrating online analytical processing, data mining and spatial analysis. The web publishing sub-system is a web site, which is squarely based on web GIS to supply e-advertisement, consultation, browsing, download and subscription and other online services.

3.2 Software platform selecting

Thanks to the rapid development of GIS technology, currently, a number of GIS platforms are suitable to serve our purpose. The mainstream GIS platforms around the world mainly encompass Arc/Info, MGE, MapInfo, AutoCAD Map, etc. Besides, a batch of domestic GIS platforms such as GeoStar, MapGIS and SuperMap have shown their strong potential with competitively high performance and low price.

When choosing GIS platform for a city the first factor to be taken into account is to carefully consider its specific construction objectives, existing data formats and related software platforms and financial budgets. For developed cities, such as Guangzhou and Shanghai, which aim to develop comprehensive urban census GIS, one can prefer software with abilities of openness, stability, and high performance towards complex urban data. In this case, enterprise GIS platforms such as ESRI’s products should be given a higher priority. On the contrary, small cities should seek an optimum balance between budgets and performance.

With a consideration of convenience and compatibility for running and maintenance of the system, Windows 98/NT/2000 are recommended as the operational system. Besides, the urban census GIS should be running under a C/S-based architecture. On Client side, Windows 98/NT/2000 serve as the operational system, while Windows NT Server on Server side.

Enterprise database software with high performance should be given a highest priority. Among those enterprise-level database systems (Oracle, In-
formix, Ingres and Sybase), Oracle database should be highlighted. Such a preference is not only because of its popularity in GIS field, but also because it serves as the mainstream platforms of storage and management of current census data. But at the same time, please bear in mind that their prominent performance cannot offset underlying disadvantages of high price and vulnerability of maintenance. Therefore, some small or developing cities should also consider other alternative databases such as MS SQL Server.

4 Some issues on urban census GIS designing

4.1 Appropriate data selection

Developing a top-ranking information system requires a careful selection and rational organization towards raw data containing rich spatial and attribute information. Compared with other information systems, the urban census GIS has its own characteristics, wide application prospect and integration ability with other socioeconomic data.

Developing an urban census GIS should also consider its potential application perspective and especially expansibility and compatibility for future updating. Therefore, it is extremely important to ensure the census GIS to be an open system. For example, when the customers intend to expand some new functions to the current census GIS, they can implement it conveniently. Considering the facts of Guangzhou city, we can give our priority to urban spatial data, proper historical census data, industry distribution data, etc. in the 5th National Census for the building-up of the urban census GIS.

When considering the selection of urban spatial data, all-feature digital vector maps are preferable. The principal factor of judging whether topographic maps meet the requirements of census GIS lies in whether maps can be easily integrated with the census data and meet the demands of different applications. As a fact, large-scale maps generally contain abundant geographic information, and thereby cause the difficulty of data maintenance along with the follow-up increase of mountains of raw data. Besides, the selection of suitable spatial data should also consider other factors such as the updating, maintenance and suitability of existing data and other materials.

4.2 Census GIS database designing

The geo-census database is one of the key parts in the urban census GIS. The goal of database designing is to realize the effective and efficient storage and management of all relevant data to satisfy the demands of customers. Regarding the design of databases, lots of publications are available in elaborating the strategy and methodologies of database design. This paper will give no explanation to that point. Nevertheless, a brief discussion on some issues concerning the unique characteristics of the urban census GIS, such as the method of geocoding, the organization and construction of data, etc. will be outlined.

The 5th National Census makes out a set of detailed census data-coding criterion on how to systematically code census "short", "long", death status and floating questionnaires. Under the guideline of these criterion, optical scanners have been used to capture those data, and based on that, nationwide uniform census databases will be created. To achieve that, geographic data must be integrated into census data in order to realize the effective linkage between them. Such a linkage can be achieved by systematic geocoding in an uniform urban reference system, determining the location of all census units, and by subdividing or aggregating some special census units.

This paper holds the idea that the coding results of census data can be a kind of important reference to the geocoding method, but are not suitable for direct use in the urban census GIS. To ensure the openness and expansibility of the census GIS, more attention should be paid to the time series of data (when were the data captured), data source (where and how were the data gained) and data fusion from different fields (data about industry, commerce, etc.). Therefore, the census GIS data should be coded with all those factors taken into consideration and keep dynamic relationship among data in database. Besides, the relatively logical and physical independence of database should also be considered in order to execute operations of adding, deleting
and updating data and to guarantee the high flexibility and performance of the census GIS.

High attention must be paid to the design and organization of data structure, which is invariably a core point in database design. TIGER system is a prominent apotheosis in this field. TIGER/line files, as the core of TIGER system, contains linear features, ground characteristic points and polygons [21]. The data structure of TIGER files consists of three kinds of data tables (0-dimension table, 1-dimension table and 2-dimension table) and corresponding indices (0-dimension index and 2-dimension index) that are stored in B-tree structure. All the census data tables are associated to these three kinds of data tables. Fig. 6 illustrates the fundamental structure of TIGER files.

Even according to today's standard, the efficient data organization of TIGER file is still worth learning for Chinese urban census GIS. Especially, its strategy of integrating various files by constructing topological relationship largely improves the efficiency of data analysis, processing and updating. In order to explore the practicable pattern suitable for the construction of Chinese urban census GIS, a further study should be given to object-oriented data organization, time series and other hot fields.

4.3 Construction of the census data warehouse

Data warehouse provides a kind of effective organism for data storage, converting various dispersed data into a uniform system, which makes them easy to access, integrate and analyze, and thereby ensures the openness of data analysis functions. When integrated with DM (data mining) and OLAP (on-line analysis process), data warehouse can largely improve the application efficiency of data analysis.

In order to realize the support to decision procedures more efficiently, decision support system (DSS) needs to access sufficient, heterogenous, present and historical data to extract necessary information. Even all raw data required are available, a further processing concerning with the generalization and aggregation towards specific and detailed data, which coincides with the characteristics of data warehouse, are still needed. Therefore, more efficient decision support can be achieved only by building data warehouse-based comprehensive DSS and its seamless integration with DSS [23].

The census data warehouse is based on urban compositive data. Complied with certain regulations, different kinds of data are extracted, cleaned, aggregated, and then transformed to the data warehouse [6]. Metadata, a very important conception to DW, elaborate data in DW, and support effective management, development and utilization towards relevant data. Fig. 7 illustrates the process of data warehouse building.

Fig. 6 The structure of TIGER files

![Fig. 6 The structure of TIGER files](image)

Even according to today's standard, the efficient data organization of TIGER file is still worth learning for Chinese urban census GIS. Especially, its strategy of integrating various files by constructing topological relationship largely improves the efficiency of data analysis, processing and updating. In order to explore the practicable pattern suitable for the construction of Chinese urban census GIS, a further study should be given to object-oriented data organization, time series and other hot fields.

4.3 Construction of the census data warehouse

Data warehouse provides a kind of effective organism for data storage, converting various dispersed data into a uniform system, which makes them easy to access, integrate and analyze, and thereby ensures the openness of data analysis functions. When integrated with DM (data mining) and OLAP (on-line analysis process), data warehouse can largely improve the application efficiency of data analysis.

In order to realize the support to decision procedures more efficiently, decision support system (DSS) needs to access sufficient, heterogenous, present and historical data to extract necessary information. Even all raw data required are available, a further processing concerning with the generalization and aggregation towards specific and detailed data, which coincides with the characteristics of data warehouse, are still needed. Therefore, more efficient decision support can be achieved only by building data warehouse-based comprehensive DSS and its seamless integration with DSS [23].

The census data warehouse is based on urban compositive data. Complied with certain regulations, different kinds of data are extracted, cleaned, aggregated, and then transformed to the data warehouse [6]. Metadata, a very important conception to DW, elaborate data in DW, and support effective management, development and utilization towards relevant data. Fig. 7 illustrates the process of data warehouse building.

Fig. 7 The process of data warehouse building

![Fig. 7 The process of data warehouse building](image)
database: conception model designing, logical model designing and physical model designing. As far as subject-oriented DW is concerned, it is corresponding to designs of information package, star-like graph model and physical data model.

4.4 Analysis methods for urban comprehensive data

The urban census data contain abundant information and can be applied in extensive fields by integrating with other kinds of tremendous complex data. How to extract useful information from those huge amounts of raw data, to large degree, depends on effective data analyzed. Fig. 8 and Fig. 9 illustrate analysis results of efficiency and space effectiveness index model of Guangzhou public utilities[7].

Data analysis covers the overall proposed urban census GIS architecture model. The comprehensive service sub-system and statistic transaction sub-system employ methodologies of spatial analysis and statistical analysis. The DSS, around the building-up of population data warehouse, concentrates on data analysis and processing by fully employing metadata technology, OLAP and data mining technology and comprehensive application decision theory and algorithms. The objective of bringing up the architecture mentioned above is to explore a series of practicable analysis methods for urban comprehensive data so as to fully analyze and utilize data.

The DSS also encompasses three layers: data management, data analysis and user interface. Consisting of database, data warehouse and data interface, the data managing layer is the essential part of the DSS, providing all necessary data for data analysis layer. The data analysis layer is composed of model database, method and knowledge database, OLAP and DM tools, and spatial analysis functions, each part of which tightly cooperates with one another and keeps its relatively independence at the same time. The user interface layer is a set of visual tools which implement dynamic user interaction functions ranging over problem analysis and generalization, querying analysis, management and maintenance, etc.

Fig. 8 The efficiency index of public facility in Guangzhou

Fig. 9 The space effectiveness index of middle schools in Guangzhou

5 Conclusion

In spite of rapid development and wide application of GIS, the construction of urban census GIS in China is still novel and challenging along with unprecedented technical and socioeconomic problems. Undoubtedly, sophisticated experience and technologies from developed countries can be used for reference, but these experiences can not fully match Chinese social and economic reality. At present, by combining with the 5th National Census, many cities are researching for the feasibility and strategies of developing Chinese urban census GIS. A further cooperation and communication among cities should be strengthened so as to construct first class Chinese urban census GIS and serve our society better. (Continued on Page 42)