Shen J, Chu D, Zhang Q and Zhang W, 1999, Developing a Census Data System in China, *International Statistical Review*, 67(2), 173-186

++

Developing a Census Data System in China

Jianfa Shen[1], David Chu[1], Qingpu Zhang[2], Weimin Zhang[3]

[1]Department of Geography, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

[2]Chinese Academy of Surveying and Mapping, 16 Beitaiping Rd., Haidian District, Beijing 100039, China

[3]State Statistical Bureau, No. 75 Yuetan Nanjie, Sanlihe, Beijing 100826, China

Email: jianfa@cuhk.edu.hk

Summary

China has conducted four population censuses since 1949. A large amount of important information about population, education, employment, migration and urbanization was collected in the most recent 1990 census. This paper will examine main features and key issues of the Chinese population census and the census data. Some fundamental considerations of building a computerized census data system and the ways that such a system might be developed will be discussed. The main objectives and features of the on-going Population GIS of China project will also be examined.

Key words: Census; Census data system; China; Population GIS.

1 Introduction

Census data have become a valuable data resource for monitoring demographic, socio-economic trends and for policy formulation, resource allocation and planning (Openshaw, 1995). This paper is concerned with the development of census data and their potential wide applications in China. The paper is focused on the development of an easily accessible computerized census data system and the main issues in developing such a system. This of course is not an easy task for a country with over 1.2 billion populations. But developments in computing technology, computer network and GIS in the 1990s have paved the way for developing such census data systems (Rees and Shen, 1997).

The paper is organized as follows. Section 2 will discuss the applications of census data in various areas such as government, academic communities and the business sector. Section 3 will focus on the most recent 1990 population census in China and its major features and limitations. Section 4 will examine the computer processing of China's 1990 census data and how an improved census data system might be developed. The Population GIS of China project will also be examined. Some conclusions are reached in section 5.

2 Making use of census data

Population census was used initially to count the number of population at a specific time in particular country. Modern population censuses have greatly extended the census content to identify the varieties of people and their socio-economic status, quality and activities across space. In recent years, with the rapid increase in computing power and analytical techniques, the census data have been increasingly used by government, academic communities and the business sector.

Government is generally assumed the task of providing at least minimum living standards for its citizens, using and allocating the public resources efficiently and effectively, and finally planning the future development. All these government activities can make use of census data which provide detailed demographic, socio-economic information about residents living in various areas. National or local government can target their resources to specific areas using detailed geo-coded and areal census data. In the case of China, population census has been used as an important means of monitoring its still growing population and the implementation of family planning programmes.

Academic communities have been involved in analyzing the changing demographic, socio-economic trends at the international, national, regional and local scales with the aim of advising the international organization, government and the general public on the challenging issues facing the society. Again, academic communities use various methods to collect data from various sources and census data are perhaps a valuable, complementary and economic data source. Various demographic, socio-economic issues, particularly at regional and local scales, can be examined using the census data.

Generally, national statistics authorities collect and publish many statistical data for the country as a whole and less statistical information is available at regional and local scales. As a result, census data for local areas turn out to be an important data resource for regional and local level studies.

Using census data for demographic research is one important application of census data. Population status in various areas can be analyzed by age, gender and ethnic groups. Population migration analysis, national and regional population modelling and projections can be conducted using the census data and other data

sources. For example, the 1982 Census data and the 1987 one percent sampling data of China have been used to make a multiregional population projection for urban-rural and provincial populations of China (Shen, 1996; Shen and Spence, 1996; 1997). The 1990 census data of China has been widely used for analyzing population change and regional migration in China (Wang, 1992; Census Office and DPS, 1993).

Census data can also be widely used for socio-economic research including issues on labour force and employment change, trends of social stratification based on social class and socio-economic groups and pattern of social life on housing, family composition and education.

There have also been increasing applications of census data in the business sector including business planning and marketing (Longley and Clarke, 1995). The most important use is perhaps in the field so-called as Geodemographics (Batey and Brown, 1995; Birkin, 1995; Goss, 1995). Geodemographics is essentially refer to the process and the classification of small scale areas. Various data sources may be used. But the most important source of information is the census data. Geodemographics systems provide a spatial dimension for companies to target their customers in addition to conventional variables such as age and social class. Potential marketing areas can be identified by linking sales performance with area types in a geodemographics system. The social-economic characteristics and consuming behaviour of customers can also be identified by linking customers' addresses with area units in a geodemographics system.

Previous discussion mentioned a wide range of applications of census data in various countries. Some applications have been made using China's census data particularly for population research and government planning. The applications in business sector are rare indicating huge potentials in this area. Nevertheless, it is

argued that a computerized census data system will greatly stimulate census data applications which will be most useful in identifying social, economical, regional and geographical problems and promoting socio-economic development in China. It will be useful to examine the status of the census in China before considering the issues of developing a geo-coded census data system in China.

3 The main features of the 1990 population census in China

Four population censuses have been carried out in China since the foundation of the People's Republic. The first and second censuses were undertaken in 1953 and 1964 respectively. These two censuses aimed to examine the population situations facing the country after its foundation and the socio-economic crisis in 1959-1961 respectively. The data on these two censuses were kept secret and were not released until recently. The 1953 census employed more than 2.5 million census enumerators. The total population in mainland China (excluding Taiwan) was 582.6 million at the time of census (Mid-year 1953) including 8.4 million population which were indirectly counted using the local government data (DPS, 1991). The 1964 census reported that there were 694.6 million population in mainland China at the mid-year 1964.

The third census was conducted in 1982 with assistance from the United Nations after economic reform and open-door policies were introduced in 1978. The results of the 1982 census were released to the outside world immediately after its completion. According to this census, the population in mainland China was 1008.2 million at mid-year 1982. Only 28601 persons in Xizhan were counted indirectly in the 1982 census.

The 1990 census was the fourth and the most recent census in China which is the main focus of this section. The relevant census procedures, the census geography, the definitions of resident population and migration population will be examined here. These are related to the ways that China's census data may be used, interpreted and the next census may be improved.

The 1990 census was organized and implemented by the state council and various local governments in China. Census leadership teams and their offices were set up at state council and local governments at provincial, prefecture and county-levels. Town and township governments and street administration branches in cities set up census offices. Census teams were set up in each villagers committee and residents committee. These organizations are in parallel with the administrative system in China. About seven million census enumerators were recruited for the census of over a billion population.

The basic area unit in the 1990 census was the census area which was equivalent to the administrative area of a villagers committee or a residents committee. The census area were further divided into census enumeration districts and each census enumerator was assigned to one enumeration district (Figure 1). The number of area units at various levels are related to the size of census data. In 1995, There were 30 provincial level units, 334 prefecture-level units, 2143 county-level units. There were also 640 cities in China, 3 at provincial level, 210 at prefecture-level and 427 at county-level and these cities administrated 706 urban districts (SSB, 1996). In rural China, there were 47,136 township-level units and 740,150 villagers committees in 1995. Thus it is clear that the census enumeration districts are arbitrary areas with no distinctive geographical meaning and no equivalent administrative unit. It is very likely that such enumeration districts will be different between censuses.

Thus there may be great difficulty in acquiring the geography boundary data and making meaningful analysis at the enumeration district level. On the other hand, census data for census areas and the administrative areas above the census area level should be readily available in paper or electronic form. The question is in what form the census data will be available. In the 1990 census, it was required that all field work should be completed over the period July 1 to July 10, 1990 and the quality checking of the census data be completed by July 15, 1990. Similar to the 1982 census, main population figures were first summed up manually to achieve an early release of the census results. Then the census forms were coded by census enumerators (coding of simple items) and 8673 data coding workers in county-level units (coding of complex items) and were input into computers. Finally, all census data were processed by computer centers at prefecture-level, provincial level and national level computing units. Electronic tabulations of the census data were produced and published selectively by various census offices. Thus in theory, raw census data for each individual and tabulated census data exist in electronic form. The problem is how to establish a common and standard electronic data sets which cover a wide range of cross-tabulated census information required by various government, academic community and the business sector. This issue will be further elaborated in the next section.

It will be interesting now to see what kind of data were collected in the 1990 census. Generally, China's census focused mainly on the demographic aspects of the population though information on education, employment status and reason for migration were also collected. Comparing with the censuses in Hong Kong and United Kingdom, the 1990 Census in China collected extra information on deaths and births which is very important for monitoring China's population trends. On the other

hand, no information on income, housing status and commuting was collected in the 1990 census. In summary, the 1990 census of China collected 15 items of information for each individual which are as follows:

1. Name;
2. Relationship with the household head;
3. Gender;
4. Age;
5. Nation;
6. Status and nature of household registration;
7. Usual residence on July 1, 1985;
8. Reason for migration to current area;
9. Education level;
10. Industry of employment;
11. Occupation;
12. Status of inactive and unemployed population;
13. Marriage status;
14. Fertility of women and number of their survived children;
15. Fertility status since January 1, 1989.

Six items of information on households were also collected:

1. Household code;
2. Household type: family household or collective household;
3. Number of population;
4. Number of births;

5. Number of deaths;

6. Number of members of the household who had left the county or city for more than one year.

China has a peculiar residence registration system in which population are registered either as agricultural population or non-agricultural population. The transfer of household registration from one place to another and from agricultural population to non-agricultural population needs an official approval. Flexible migration policies have been introduced in China so that migrants are allowed to register temporary or permanently at their destinations since the late 1970s. But the formal transfer of household registration still lags behind the actual population migration. Thus there are increasing number of population whose place of household registration is not the place of usual residence. Thus the 1990 census cannot use the household registration to define usual residents. Instead, a special definition is used which include five types of population as the census population in a county or city:

1. Those living in the county or city and their household registrations were also in this county or city;

2. Those who had lived in the county or city for more than one year but their household registrations were in another county or city;

3. Those who had lived in the county or city for less than one year but they had left their place (county or city) of household registration for more than a year;

4. Those who were present in the county or city whose household registration needs to be determined;

5. Those who originally lived in the county or city but had moved to overseas to work or study thus had no household registration.

Thus the definition of census population in the 1990 census is different from the population with formal household registration in a county or city. Those individuals who had left their place of household registration for over a year were counted as census population in their current place of living. The concept of census population is important as it is the base population for any further cross-tabulation and analysis. There are two potential problems in the 1990 census definition.

First, not all persons who had lived in a county or city for less than a year were counted in the census population. If person A, person B and person C arrived in a particular city at the same time and had spend six months by the time of census, person A was counted in the census population at destination if his household registration had been transferred, person B was not counted in the census population in the destination if he moved directly from his origin county while person C was counted in the census population in the destination if he left his origin county one year ago and had spent some time in another city or county.

Second, it appeared that population with different household registration and resident place but within the same county or city were counted in the census population of their place of household registration. Thus local level population change under the county-level may not be detected by the 1990 census though it is very likely that there are large volumes of population movement at local scale. Both the gravity model and evidences in other countries indicate that there will be increasing number of migrations between origin and destinations as the distance between them becomes shorter (Stillwell and Congdon, 1991; Boyle, Flowerdew and Shen, 1995).

The concepts of the urban population and rural population are also important in interpreting the 1990 census data. Generally, population figures for three types of residence including street, town and township are available. However, a city or a county may include all three types of residence and both agricultural population and non-agricultural population may register in towns. In the 1982 census, all populations in cities and towns are counted as urban population. But this definition is no longer useful as many counties and townships have been designated as a city or a town respectively. The proportion of urban population may be more than 50% if the 1982 census definition of urban population was adopted in the 1990 census. In the 1990 Census, the State Statistical Bureau of China adopted another definition of the urban population. Cities were divided into two types according to whether a city was further divided into city districts or not. All persons, including the registered 'agricultural' and 'non-agricultural' populations, registered either with residents committees or with villagers committees, in city districts were counted as urban. This is in line with the 1982 definition. Only persons registered with residents committees in cities which did not have city districts and in towns were counted as urban. The persons registered with residents committees included all registered 'non-agricultural population' and a small part of 'agricultural population' who had moved into urban areas and was allowed to register with residents committees. The overall size of urban population based on the 1990 census may be realistic but the figures between different spatial units may not be perfectly comparable. About 65 million out of the 187 million urban population in city districts were registered as 'agricultural population' (those who registered with villagers committees) while none of the 'agricultural population' registered with villagers committees were included in the 109 million urban population in the remaining urban areas (DPS, 1991). Thus comparison of the urban

population between a county and a city may not have a consistent basis. Detailed computerized census data may help to overcome this problem by allowing users to derive appropriate population base.

Another important concept in the 1990 census is the definition of migration. One question in the 1990 census asked about whether the usual residence on July 1, 1985 (five years before the census) was the same as the current county or city and, if different, whether it was a county or city in another province. Thus from the census form, only the specific province of origin can be identified, the origins of intra-province migration can not be identified and no information was collected about the intra-county or intra-city migration. Migrants who have moved into a destination for just one year and up to five years will be treated as the same. In other words, the migration data from the 1990 census is very limited in terms of both the time and spatial scale. Slight change in the census form will generate much more useful migration information. As mentioned before, it is at the local level that majority of population migrations take place .

Three types of residence place were identified for inter-county or inter-province migrations for the period 1985-1990. Thus migration flows among the residence types street, town and township can be tabulated. However, the residence type of a place at the beginning of the period was based on the status of the place at that time instead of the status at the end of the period. Thus if there was a migration flow between town A and town B based on the status in 1990, the flow would be recorded as a township to town flow in the 1990 census data if the status of town A was a township in 1985. Here there is a problem of transition of the place type in the migration period 1985-1990. It is unclear as whether the place type at the beginning or the end of period should be used. The migration event and the transition event of

the place status could happen at any time between 1985 and 1990. Theoretically, if migration takes place after the transition of the place status, the flow should be recoded using the place status at the end of period. On the other hand, if migration takes place before the transition of the place status, the flow should be recorded using the place status at the beginning of the period. Figure 2 presents these two situations. The definition used in the 1990 census is just one of several possibilities.

4 Developing census data system in China

The previous section examined the major features and issues of the 1990 census data in China. The next census will take place in the year 2000. There are still some scope in improving the dissemination of the 1990 census data and some new approaches might be adopted in the design, implementation, computer processing and dissemination of the 2000 census. This section will discuss the ways that a computerized census data system might be implemented in China.

Generally speaking, computerized or electronic form of census data system can be established using simple to more sophisticated techniques from computer disks, CD-ROM to networked access to the census data. It might appear to be very difficult to build a computerized census data system for over 1.2 billion population in China. However, it will not be so difficult if we see the census data system as a system consisting of several independent parts. The whole system might be implemented in a step by step manner. For tabulated census data, the size of census data is determined by the spatial units involved. For example, there were 47,136 township-level units and 740,150 villagers committees in rural China in 1995.

It will be interesting to examine the implementation of computer processing of the 1990 census data (Wang and Yu, 1993; Teng, 1993; Li and Lan, 1993). For a

large country like China, the total census data involved is enormous. It was estimated that the total raw census data amounted to 60 GB in 1990. At the county-level, the maximum number of population was 2 million and the volume of raw data was 100 MB. At the prefecture-level, the maximum number of population was 14 million and the volume of raw data was 700 MB. At the provincial level, the maximum number of population was 107 million and the volume of raw data was 5350 MB. The 1990 census data were processed in two phases. Only 10% sampling data were processed in the first phase for earlier delivery of the census results. All data were processed in the second phase to produce 100% census tabulations. The first phase was completed by December 1990 and the second phase was completed by April 1992.

The total number of 100% census tables were 62 at county-level, 419 at prefecture-level, 710 at provincial level and 852 at the national level. The size of data volume varies at different levels. For 100% provincial census tables, the sizes of tabulated census data were 50 MB without printing format and 250 MB with printing format respectively. The procedures to extract these census tabulations from the raw census data depend on the computer hardware and software available in the 1990 census. The 1990 census data were processed in four stages at three level as shown in figure 3 (Wang and Yu, 1993). As mentioned before, initial data input was carried out at county-level computing units with a few exceptions. The raw census data for each township-level units were checked and edited in the first stage of processing at the prefecture-level computing units. In the second stage at the prefecture-level computing units, the township-level data files were merged to generate county data files which were used to generate county-level census data tabulations. It is noted that the county data file was no longer a simple list of individual records, but have been tabulated and re-coded to save data storage space. This has been necessary due to the

limitation of computing power. On average, the size of the total raw census records for a county was 50 MB and the size of tabulated census data at provincial level and national level was also about 50 MB. Direct processing of either the raw census records or the fully tabulated county-level data at the provincial and national level was not feasible at that time. In the third stage which was carried out at provincial level computing units, census tables for each provincial units were tabulated and the national tabulations were carried out at the national level in the fourth stage.

It is clear that the 1990 census data were processed in a decentralized manner and raw census data were pre-processed before submitting to the higher-level computing units for tabulation of higher spatial level census data. Different number of census data tables were produced for different spatial levels. In recent years, there has been great advance in computing hardware and software. Now it is not too difficult or expensive to store tens or hundreds of MBs of raw census data or tabulated census data in one piece of electronic media and to be easily transferred. The implementation of computer network may made it possible to transfer all census raw data to the national center for central processing and dissemination. Thus in the next China census, new mode of census data processing and dissemination may be adopted with improved computing facilities.

Figure 4 shows the key stages of developing a census data system and its main components. Some components have already been in operation such as the current computer storage of the raw census forms and the tabulation of current census results. Some components need to be expanded/standardized/improved. Some components need to be developed from fresh such as the digitized boundary data for spatial units under county level and the networked access to the census data system. The key components and implementing stages of a census data system is discussed below.

1. The design of census form. The census results depend on what questions are asked and what kind of detailed information is to be collected. The 1990 census has been completed and nothing can be done about that. Some considerations might be useful for the design of the 2000 census. It is always a good practice to invite census users from government, business and academic sectors to discuss their needs for census data.

In the 2000 census, it may be appropriate to consider including a few more questions in the census form on income (by income groups), housing (type, tenure, number of rooms and space), mode and time of travel to work. For migration question, it is highly desirably to record the place of origin down to the county or township/town/street level. In the 1990 census, only the origin province of inter-provincial migration can be identified and is not sufficient for detailed socio-economic analysis related to migration. To reduce the coding workload in the 2000 census, only the origin town/township may be coded for intra-county migration and origin county be coded for all other inter-county and inter-provincial migration. Another alternative is only to introduce these additional questions for more developed provinces in China.

Indeed, some progress has been made in the 1% population sampling survey conducted in October, 1995 (Zhang and Yu, 1995). New questions on housing were included and several questions on usual residence, migration, education and unemployment were improved. The geographical area to define usual population was reduced from county-level units in the 1990 census to township-level units in the 1995 sampling survey. Inter-county migration data may also be derived from the 1995

sampling data. These will greatly enhance the value of population data. It is expected that these and other improvements will be introduced in the 2000 census.

2. Raw census data. In the 1990 census, all census forms were inputted into the computer. Thus a set of computerized census raw data actually exist in electronic form. This data set can be used to generate more detailed census data than those that have been tabulated and published. In the future, with the availability of more powerful computers and workstations, raw census data inputted in the local level can be sent to provincial and/or national level computing center for data processing and tabulations. In the initial stages, these raw data can be transferred by disks and/or high capacity storage devices. The computer network can also be used to transfer all or part of raw census data when such a computer network is completed in China.

3. Census data sets. The derivable census data from the census raw data depend on the original census form. What is needed is a set of standard census tables and related software to extract these data from the raw census data and these standard data then can be made available in a standard format. Such a date set may be called Regional Statistics as it contains census data for various geographical units. Special software may be developed for users to extract a sub-set of theses census data. For example, some 100 standard census tables may be designed for the 1990 census data set. According to the census geography mentioned before, the same set of data will be available for China as a whole, provincial units, county-units, town and township units. These data may be organized into a number of files. For example, one file may include all data for provincial units in China. Each province may have a file including all information for all county-units in the province. Each county may have a file including all information for all township/town/street units. The most important of all is that these data should have a standard structure, the users can make use of such

large number of files very easily. In other words, there will be a common data sets for all spatial units at various levels. Users will only need to know the area name/code and the census variable name/code to extract the required information from the census data system. Right now, various county-level census results have been published with different coverage causing difficulty in using census data from various counties.

Another set of data called Migration Statistics might be extracted from the 1990 census. This Migration Statistics may consist of two sub-sets of data. Migration Statistics One would provide detailed tabulations according to each inter-provincial flows and other demographic, socio-economic characteristics of migrations. Migration Statistics Two is about intra-provincial migrations and detailed information can be tabulated for two categories of migration in the 1990 census: between-county migrants and within county migrants.

Finally, it is worth considering to provide a set of anonymized samples with individual data so that users can define new variables and derive new cross-tabulations. Such data will allow the application of a multi-level modelling approach to avoid model mis-specification using only conventional individual level or aggregated areal data (Boyle and Shen, 1997). This data set may be called Microdata. For a large country like China, the Microdata may be provided as 1% sample files for each provincial unit and the lowest identifiable geographical unit may be the county to reduce the exposure risk of individual identities. In the 1990 census, only the origin province of migrants will be available in the Microdata.

4. Census data extraction software. The extended census data sets can be provided in plain format or in a special format with extracting software allowing users to extract variables for specific area units on PC or Workstation platforms.

5. The digital boundary data. It is highly desirable that digitized boundary data can be provided for the whole census geography down to the streets/township/town level. This will allow efficient analysis of the census data using GIS and other mapping software. Indeed, the provincial level and county-level boundary data have already been produced and these data can be made available to other users in a suitable format such as Arc/Info GENERATE format (Liu, 1995). The main workload will be on the development of boundary data for township-level units and even the villagers committee/residents committee-level. But this can be carried out on a basis of county by county, city by city or province by province basis. It would be rare to use the whole set of national digitized boundary data at such a lower level. Majority of the use of such data may be for a particular county/city/province. But it is important to have a standard system for the digitized boundary data so that all these data are consistent and can be linked together easily. Other geographical data may be added to these digitized boundary data when resources are available.

6. Media of census data sets and digitized boundary data. CD-ROM may be considered as a suitable format to distribute these census data, a suitable software could be provided for easy extraction of the required census data. This could be a practical option at this stage. A number of CD-ROMs include various data sets with relevant data extraction software could be produced for the country as a whole with provincial/county level data and for each provinces with county/township level data. There is a large demand for such kind of census data and it is economical to produce CD-ROM on a large scale. For the census data for each county at a the lowest spatial scale such as the census areas, it may be more economical to provide the data on high capacity drives or tapes as the demand for copies of such data may be much less. Nevertheless, such kind of local area census data is valuable for urban and regional

planning and specialized studies. It is suggested that the State Statistical Bureau or a third party should set up a central office for producing and distribution of the census and relevant digital boundary data. The central office could be funded partly by the central government and partly by the users of census data. Different prices could be specified for central and local governments, higher education/research community and the business sector. In principle, the users in business sector should bear the full cost of collecting, processing and distributing the census data. Users in government and higher education/research sector may only pay the production and distribution cost of the media of the data as data collection has been funded by the government. It may be worthwhile to reach an agreement between the State Statistical Bureau, central and local governments, State Science Commission and State Education Commission about the funding and pricing of the census data to users in government, education and research sectors. Such kind of arrangement will be useful to achieve minimum cost of census data production and efficient use of census data.

Computer network is now being constructed in China. For example, China Education and Research Network (CERNET) is under construction and many universities have been connected with the Internet. Thus it may be possible to distribute census data over the computer network in future. Government, higher education and business sectors may have their own computer network or access to computer network. The construction of such computer networks should be funded by the relevant user sectors. A central census data unit may need to be established to manage the server of the census data and the users of census data may need to pay the operation cost of distributing the census data over the network. This central census data unit could be operated under the State Statistical Bureau, State Education Commission or a third party. The forms of network access to census data range from

simple download of census data file, on-line census data extraction or on-line census data visualization and analysis. As it is expensive and takes a long time to complete a nationwide computer network, provisions should be may to provide census data to users who do not have access to computer network. Service centers in main cities could be set up to deliver the census data to these users or to allow these users to go to such service centers to retrieve census data.

7. GIS, modelling and analysis software. Once fundamental and standard census data sets have been constructed, various GIS and modelling software can be employed for in-depth demographic, social and economic analysis. This will greatly boost socio-economic and geographical research on China.

8. Value added census data products. These can be produced to meet the various needs of government and business sector for planning, marketing and decision making.

It will also be worthwhile to mention the important role of scientific management in the development and dissemination of the census data system. These include pre-census consultation of census users on data requirement and census form; detailed documentation and definition of census geographical area units, census variables, tables, data sets; detailed documentation on relevant data extracting and analysis software; cost recovery and the licensing of the census data sets. There are many fundamental geographical, demographic, social and economic studies which may be carried out using census data in addition to studies on the census data processing, distribution and analysis techniques. These studies should be funded mainly by the National Foundation for Sciences Research in China.

Finally, further research is needed on the issues of changing geographical boundaries of census area units. They will affect the comparison between two or more

census periods and such dynamic analysis is also an important application area of census data. The numbers of township-level and villagers committees in rural China were reduced over the period 1991 and 1995 (SSB, 1996). Some units have been merged or re-organized while others have become part of urban area. The increasing number of towns is due to the transition of status from a township to a town. On this aspect, surface-based population modelling may be a useful technique to overcome the problem of changing geographical boundaries (Martin, 1989; 1995; Bracken and Martin, 1989)

Initial steps have been taken in China to develop census data systems based on GIS technology. A major research project called Population GIS of China has been carried out since 1995 (Zhang and Zhang, 1997). The State Statistical Bureau of China and the National Bureau of Surveying and Mapping agreed to establish the population GIS of China to meet the needs of the year 2000 census. The main aim of the project is to develop digitized spatial information including census area boundaries, physical features and human settlements and to integrate the spatial information with the census data. Topographic data base will be established which will include several layers for settlement, communications, river systems, administrative boundaries, vegetation and streets. Aerial photography will be used as the major source of spatial information.

For a huge country such as China, it is time consuming to digitize the spatial information for its whole territory. Thus for the period up to year 2000, the Population GIS of China project will focus on the following main tasks:

a. To develop a geographic coding system for the traditional system of administrative areas. It is proposed to use a five level 10 digits code to represent administrative areas in China.

b. To develop specification and standards for Population GIS of China.

c. To develop and improve software systems for spatial analysis and application of census data.

d. Population GIS will be developed for 200 major cities in China with both spatial and census data bases. The spatial data base will be based on maps at scale 1:10000.

e. Provincial level Population GIS will use the spatial data base based on maps at scale 1:50000.

f. The national level Population GIS will use the spatial data base based on maps at scale 1:250000.

Experimental Population GIS systems are being developed in a few cities in China. It seems that, at this stage, the main focus is on developing standalone GIS systems integrating both spatial information and census data for various cities and provinces respectively. However if same data format and software are used, it will be easy to share and transfer spatial data and census data between different systems. Much more effort should be made in designing the format of spatial data base and the census data sets, and developing specialized software for managing, extracting and analyzing the spatial and census data as mentioned in the previous sections.

5 Conclusion

The major features of the 1990 census of China were examined and some key issues and the limitation of the 1990 census were analyzed. It is revealed that the

major focus of the population census in China was still on the demographic dimension. Extra and improved questions on income, housing and commuting are recommended to be considered in the next 2000 census.

The computer processing of the 1990 census in China was also discussed and the main stages of developing an easily accessible computerized census data system and its major components were identified. Major issues in implementing a census data system in China were discussed with preliminary recommendations.

Initially, the main efforts should be made to design standard data format for census data sets which can be easily integrated with the spatial data base. It is important that digital boundary data for various administrative areas and relevant data extracting and analysis software were made available. The funding of providing census data to government, education and research community should be provided in principle by the government. A central office of census data may need to be established to manage and co-ordinate the provision of census data, spatial data and software to users.

The ongoing Population GIS of China project which aims to achieve this is also discussed. The project will develop population GIS for 200 major cities in China in its initial period up to the year 2000. Adopting common data format and software will enable various Population GIS systems to share the data. Experimental Population GIS systems are being developed in a few cities in China. The lessons from this experience will be useful to develop the state of art census data systems in China.

Acknowledgment

This research is funded by a Direct Research Grant of the Chinese University of Hong Kong, Project code 2020328. The helpful comments of the anonymous referees and the editor of this journal on an earlier draft of the paper are gratefully acknowledged.

References

Batey P. and Brown P. (1995). From human ecology to customer targeting: the evolution of geodemographics. in *GIS for Business and Service Planning*, Eds. P. Longley and G. Clarke, pp. 77-103. Cambridge: GeoInformation International.

Birkin M. (1995). Customer targeting, geodemographics and lifestyle approaches. in *GIS for Business and Service Planning*, Eds. P. Longley and G. Clarke, pp. 104-149. Cambridge: GeoInformation International.

Bracken I. and Martin D. (1989). The generation of spatial population distributions from census centroid data. *Environment and Planning A* **21**, 537-543.

Boyle P., Flowerdew R. and Shen J. (1995). Analysing local area age- and sex-specific migration flow data from the 1991 Census. *Migration Unit Research Paper* 9, Department of Geography, University of Wales Swansea, pp. 1-32.

Boyle P. J. and Shen J. (1997). Public housing and migration: a multi-level modelling approach. *International Journal of Population Geography* **3**, 227-242.

Census Office and DPS (Eds) (1993). The *1990 Population Census: Essays Presented in International Seminar*. Census Office of the State Council and Dept. of Population Statistics of the State Statistical Bureau. Beijing: China Statistical Publishing House.

DPS (1991). *China Population Statistics Yearbook 1990*. Dept. of Population Statistics, State Statistical Bureau. Beijing: China Statistics Press.

Goss J. (1995). Marketing the new marketing: the strategic discourse of geodemographic information systems. in*Ground Truth, the Social Implications of Geographic Information Systems*, Ed. J. Pickles, pp. 130-170. New York: Guildford Press.

Li Xue-zhi and Lan Yun-wen (1993). The national-level data processing system of the fourth national population census. in*The 1990 Population Census: Essays Presented in International Seminar*, Eds. Census Office of the State Council and Dept. of Population Statistics of the State Statistical Bureau, pp. 128-137. Beijing: China Statistical Publishing House.

Liu Yue (1995). The design of the national economic Atlas of China (NEAC) and the methods and technology of cartographic visualization. *ACTA GEOGRAPHICA SINICA* **50**, 193-205.

Longley P. and Clarke G. (Eds) (1995). *GIS for Business and Service Planning*. Cambridge: GeoInformation International.

Martin D. (1989). Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers* **14**, 90-97.

Martin D. (1995). Censuses and the modelling of population in GIS. in *GIS for Business and Service Planning*, Eds. P. Longley and G. Clarke, pp. 48-72. Cambridge: GeoInformation International.

Openshaw S. (Ed) (1995). *The Census Users Handbook*. Cambridge: GeoInformation International.

Rees P. H. and Shen J. (1997) The ESRC/JISC census programme and the census data system in UK. *Workshop on GIS in Spatial Population Analysis and Regional Economic Development*, organized by Department of Geography, The Chinese University of Hong Kong, Hong Kong, March 24-25, 1997.

Shen J. (1996). Internal Migration and Regional Population Dynamics in China. *Progress in Planning* **45**, 123-188.

Shen J. and Spence N. A. (1996). Modelling urban-rural population growth in China. *Environment and Planning A* **28**, 1417-1444.

Shen J. and Spence N. A. (1997) Modelling regional population growth in China. *Mathematical Population Studies* **6**, 241-274.

Stillwell J. C. H. and Congdon P. (Eds) (1991). *Migration Models: Macro and Micro Approaches*. London: Belhaven.

SSB (1996). *China Statistical Yearbook 1996*. State Statistical Bureau. Beijing: China Statistical Publishing House.

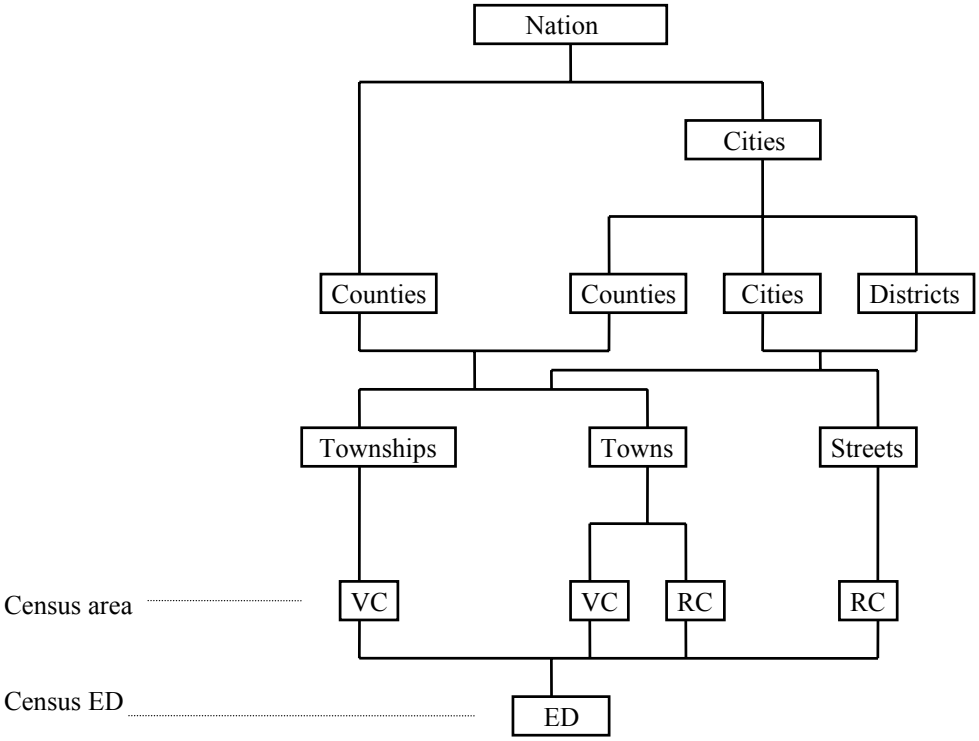
Teng Ruo-bo (1993). The design and implementation of the tabulation programme system using micro-computers for the 1990 census of China. in *The 1990 Population Census: Essays Presented in International Seminar*, Eds. Census Office of the State Council and Dept. of Population Statistics of the State Statistical Bureau, pp. 110-117. Beijing: China Statistical Publishing House.

Wang Hui and Yu Xuan (1993). Design and implementation of the computer data processing system for the 1990 census of China. in *The 1990 Population Census: Essays Presented in International Seminar*, Eds. Census Office of the State Council and Dept. of Population Statistics of the State Statistical Bureau, pp. 91-109. Beijing: China Statistical Publishing House.

Wang G. (1992). Analyzing the relationship between the inter-provincial migration to Shanghai and distance. *Population Research* **16**(4), 1-7.

Zhang W. and Yu H. (1995). The 1995 1% population sampling survey, *Population Research* **19**(3), 15-18.

Zhang Q. and Zhang W. (1997). Establishment of Population GIS of China *Workshop on GIS in Spatial Population Analysis and Regional Economic Development*, organized by Department of Geography, The Chinese University of Hong Kong, Hong Kong, March 24-25, 1997.



VC: villagers committee
RC: residents committee
ED: Enumeration district

Figure 1. The census geography in the 1990 census of China

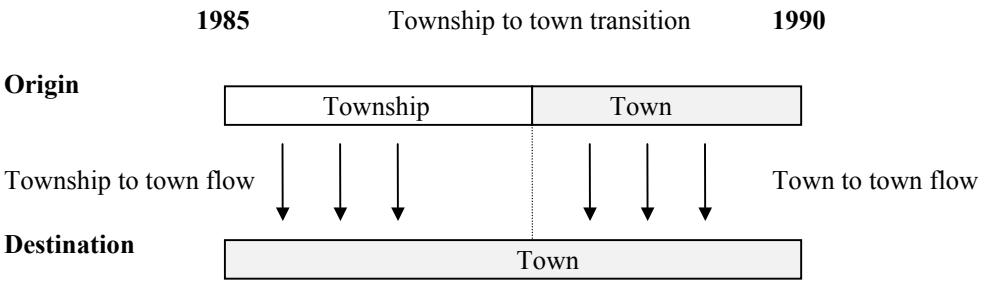


Figure 2. Defining migration flows between township and town residence types

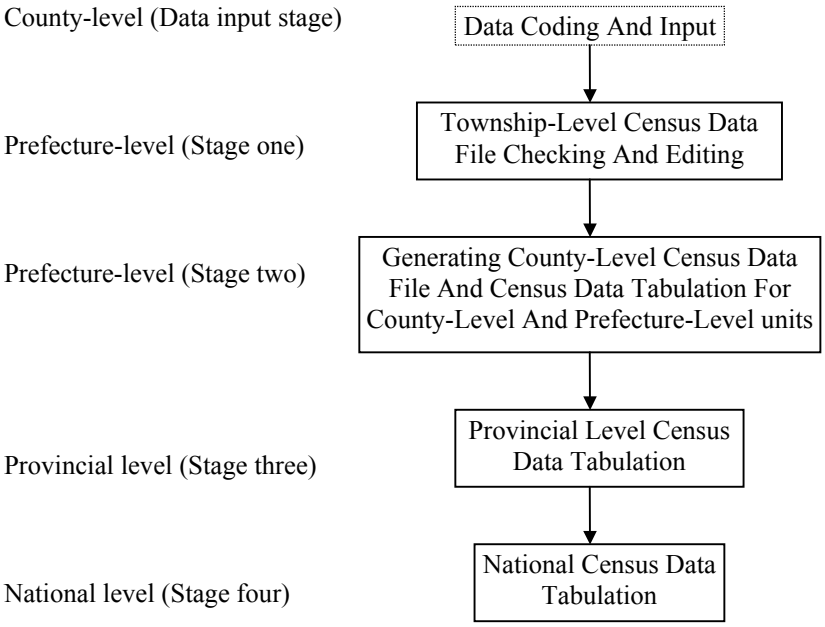


Figure 3. The four data processing stages of the 1990 census at three levels

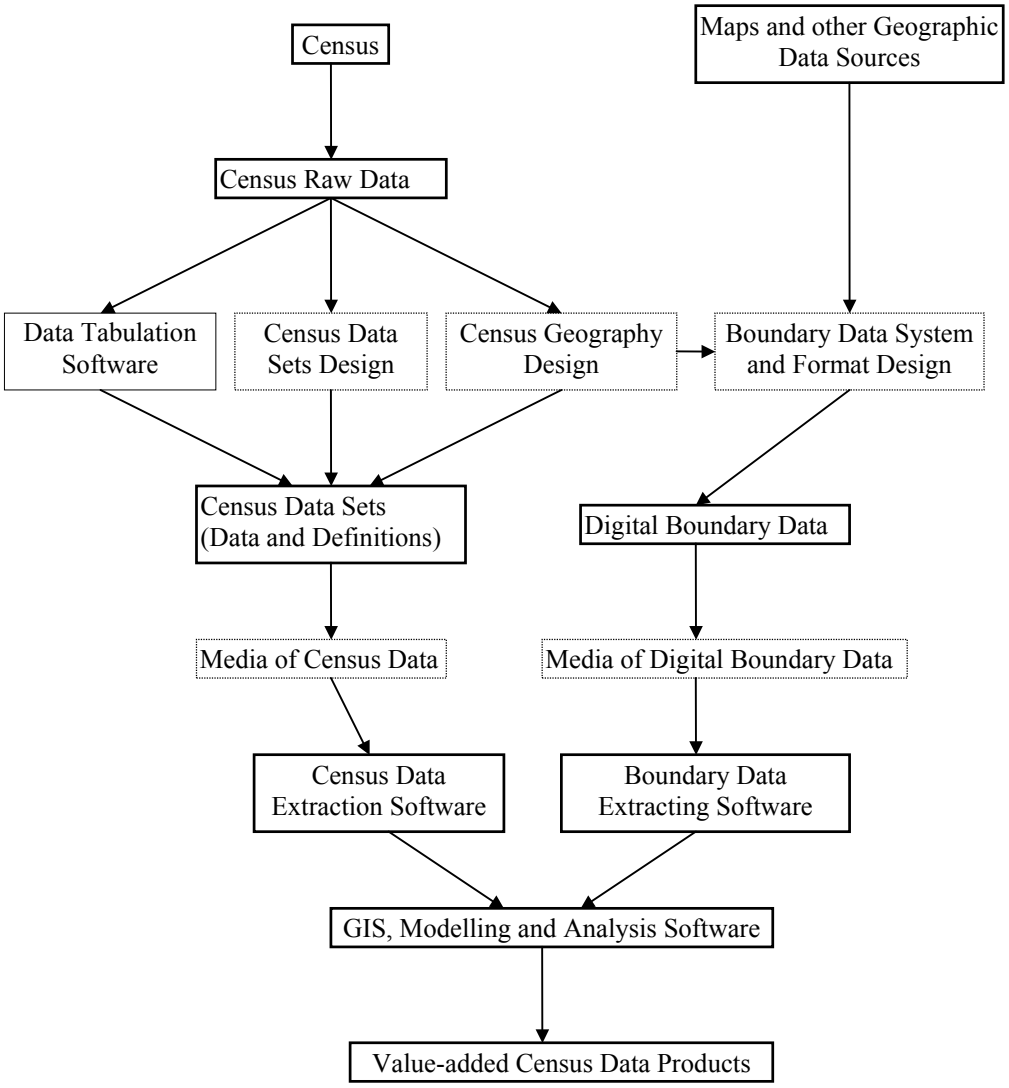


Figure 4. Key development stages and components of census data system