

Improvements of the Census Operation of Japan by Using Information Technology

By

Naoki Kurihara

Statistics Bureau, Japan

1. Introduction

- The Population Census of Japan has been conducted by the Statistics Bureau every five years since 1920, and the coming Census is to be taken as of 1 October 2005.
- With the population of Japan over 120 million, the Census is the largest statistical undertaking in Japan.
- To collect and process the extremely large amount of data within a limited time, the use of information technology (I T) is essential.

- In the recent Censuses of Japan, major improvements by the use of IT have been brought about in the following areas.
 - a. Mapping of enumeration districts
 - b. Data capture from questionnaires to computer
 - c. Dissemination of small area statistics
- The so-called Internet enumeration will not be adopted in the 2005 Census.
- However, IT will be utilized to the extent possible in the management of the Census operation.

2. Mapping of Enumeration Districts

- **Field work by the local governments**
 - The field work to enumerate the population is entrusted to the local governments, i.e. 47 prefectures and about 2,500 municipalities.
 - The work of demarcation and mapping of Enumeration Districts (EDs) was computerized for the first time in the 1995 Census on a limited scale.
 - Before the 1995 Census, local governments had to produce ED maps by cutting and pasting paper maps.

● **The use of GIS for mapping of EDs**

- The geographic information system (GIS) is used for demarcating EDs and producing ED maps.
- For computerized mapping, the digital data of the ED boundaries and digital base maps are needed.
- The digital data of ED boundaries owe very much to the Census Mapping System (CMS) which the Statistics Bureau develop by applying GIS in the 1990 Census.
- As for digital base maps, as they have become increasingly popular for PCs and car navigation systems, their prices have become more affordable.

● **Support of the local governments by providing tools**

- For the 2005 Census, the Statistics Bureau has developed a new system of computerized mapping of EDs.
- The system can produce ED maps by putting digital data of ED boundaries on the digital base maps.
- The Statistics Bureau has provided local governments (prefectures) with the system along with the digital data of the previous ED boundaries and the digital base maps.
- The coverage of such ED maps produced by computer has increased up to approximately 80% of all the EDs in Japan.

3. Data Capture from Questionnaires to Computer

- **Data capture operation by the Statistics Center**
 - In Japan, data capture operation is centrally done by the Statistics Center.
 - The Statistics Center used optical mark readers (OMRs) for data capture since 1965 until the 1995 Census.
 - In the 2000 Census, the optical character readers (OCRs) were used for the first time because of the following advantages over OMRs:
 - (a) Capable of capturing numerical responses directly as well as marks;
 - (b) Capable of capturing image data on a full scale.

● **Advantages of OCRs**

- Owing to the first advantage, it became possible to design a user-friendlier and compact questionnaire in the 2000 Census.
- The second advantage has made the data editing work more efficient because the editing staff can refer to the full image data of the questionnaire on the screen of PC.
- With OCRs, necessary information can be retrieved on the spot electronically. This change has enabled the staff to make faster and better judgment.

● **Recognition problems**

- OCRs, however, are not free from problems. As for accuracy, in some cases, non-recognition or false recognition occurs.
- The recognition performance of OCRs depends on not only the ability of the machine but also the quality of written characters.
- The speed of recognition also varies according to the complexity of the forms and the legibility of characters.
- For the purpose of the Census of Japan, the character set has been limited to the numbers so that the recognition rate and the speed can increase.

- **Coping with the recognition problems (1)**
 - To cope with the recognition problem, great care is taken in every phase of work from planning to implementation.
 - In the planning phase, OCR models are first selected on the basis of catalog specifications.
 - Then, the machines are tested with the real questionnaires of the pilot surveys to measure the performance, i.e. accuracy and speed.
 - According to the result of the test, the patterns of handwriting that are unrecognizable or falsely recognized are analyzed.

● **Coping with the recognition problems (2)**

- In the implementation phase, unrecognized characters are judged and entered manually.
- If non-recognition occurs, the unrecognizable characters are displayed on the operator's screen, and he/she judges it and enters the correct numbers.
- Moreover, there is another group of operators assigned to monitor the accuracy of recognition. Sample inspection is applied to the batches of questionnaires.
- As a result of the control, the non-recognition rate was 0.8%, and the false recognition rate was estimated to be 0.1% at maximum for the 2000 Census.

● **The possibility of capturing hand-written characters**

- In the 2005 Census, the possibility of capturing the Japanese characters was tested to introduce automated coding of hand-written responses for the first time in the Census.
- The test was targeted to the coding of “destination of commuting”.
- The result of the test showed that the accuracy of recognition was not sufficiently high, and it was concluded that automated coding of hand-written responses will not be adopted in the 2005 Census.

4. Dissemination of Small Area Statistics

● The Statistical GIS Plaza

- In February 2004, a new web site named “Statistical GIS Plaza” was launched to provide small area statistics of the 2000 Census.
- The small area statistics provided in the site are compiled on the basis of geographic units called “cho” or “aza” (address blocks).
- For each address blocks, statistics of such basic characteristics as population by age group, industry and occupation are available.

● **Various functions of the site**

- The statistics for area blocks are displayed on the base map showing the boundaries of area blocks and geographical objects such as roads, railways, rivers, etc.
- The site provides various functions to sum up statistics of two or more address blocks.
- For example, the population of an area within a certain radius from an arbitrary point can be easily computed.

➤ Statistical GIS Plaza

a.stat.go.jp/GISPlaza/ 移動 リンク

StatisticsBureau 総務省 統計局 統計 GIS プラザ 統計データ・ポータルサイト ~政府統計の総合窓口~

トップページ 操作解説 用語解説 合併情報 合併情報の取扱 収録データ ダウンロードデータ

町丁・字選択 地図操作 離計測 印刷 グラフ 選択字 ラベル レイヤ透明度 表示地域変

集計項目選択 グラフ設定 ラベル設定

集計項目 グラフ凡例 背景凡例

統計表切替 選択字表示 集計モード

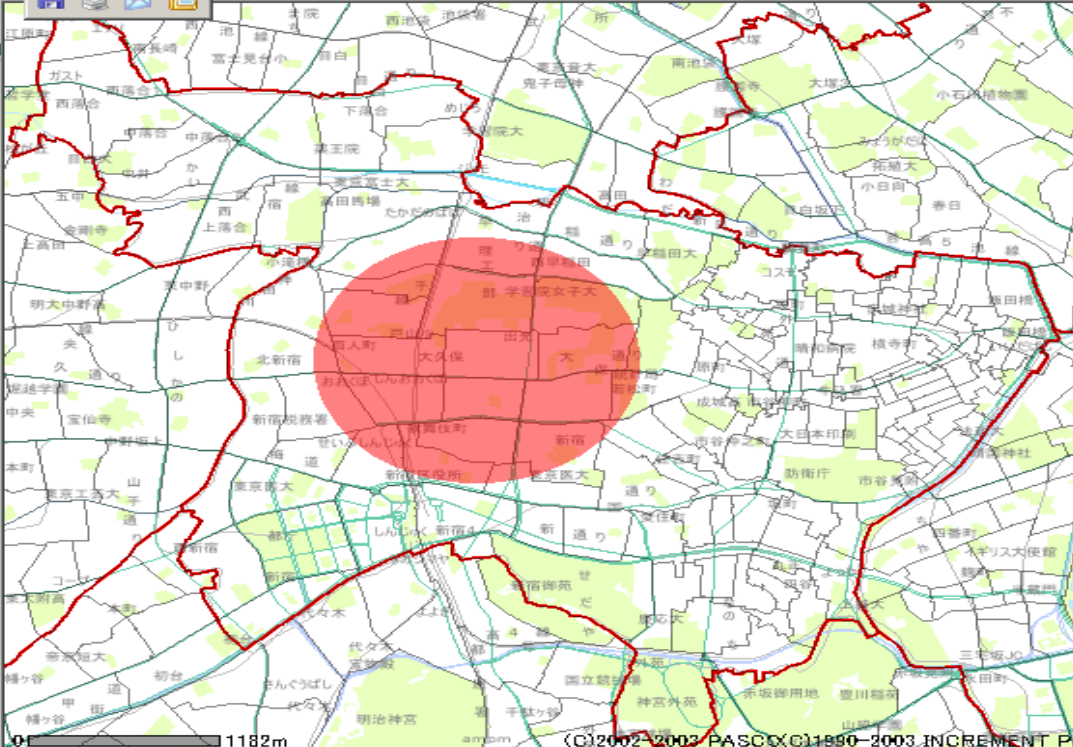
平成12年国勢調査(人口)

項目	値	単位
人口総数	0	人
男	0	人
女	0	人
世帯総数	0	世帯

同心円選択 - Microsoft Internet ...

同心円選択

半径 メートル



0 1182m

©2002-2003 PASCO ©1990-2003 INCREMENT P

52° 経度 139° 41' 49" 世界測地系緯度 35° 41' 04" 経度 139° 41' 38" インターネット

<http://gisplaza.stat.go.jp/GISPlaza/>

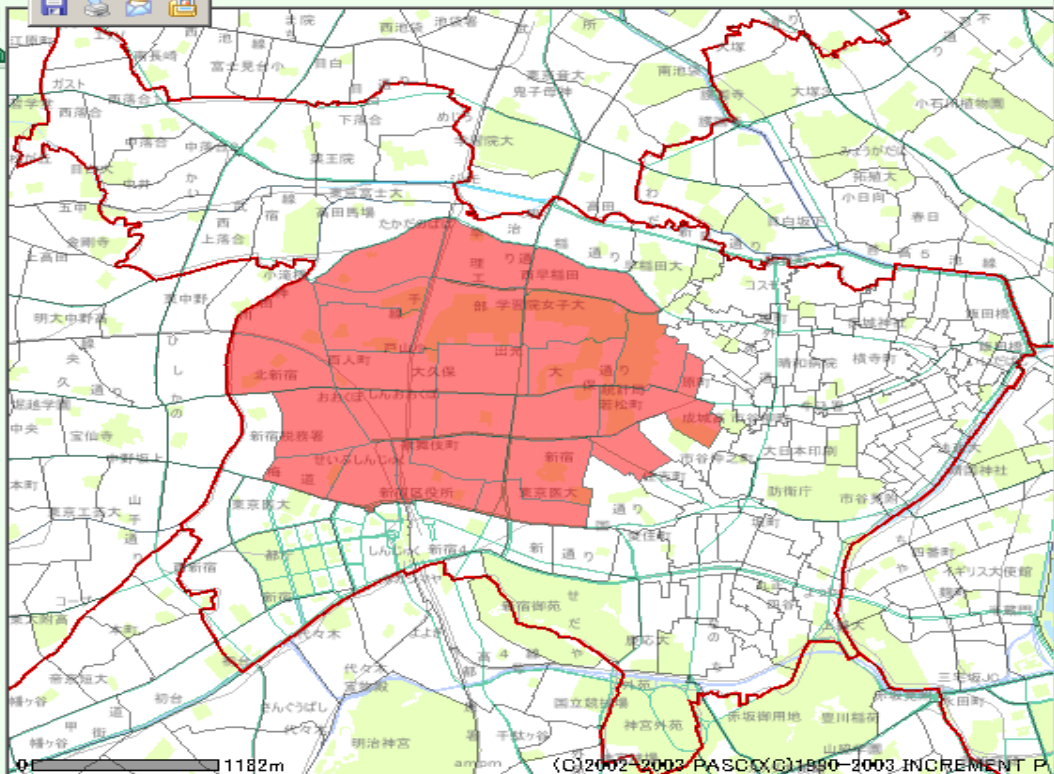
トップページ 操作解説 用語解説 合併情報 合併情報の取扱 収録データ ダウンロードデータ

町丁・字選択 地図操作 距離計測 印刷 グラフ 選択字 ラベル レイヤ透明度 表示地域変

集計項目選択 グラフ設定 ラベル設定
集計項目 グラフ凡例 背景凡例
統計表切換 選択字表示 集計モード

平成12年国勢調査(人口)

項目	値	単位
人口総数	95637	人
男	48273	人
女	47364	人
世帯総数	52719	世帯



(C)2002-2003 PASCO(C)1990-2003 INCREMENT P

● **Protecting confidentiality**

- In designing this site, special attention was paid to the privacy concern.
- As the area unit is small, there is a risk of disclosing an individual's characteristics inadvertently.
- Therefore, the statistics provided in the Statistical GIS Plaza have been limited to the basic characteristics without cross-classifications.
- The balance of wider use of statistics and protection of privacy will remain to be an important issue in disseminating small area statistics.

● **Improvements of the web site of the Census Results**

- Besides the Statistical GIS Plaza, the Statistics Bureau runs the web site of the Census results, which have more than 600,000 accesses every month.
- People can get most of the Census tables through the site free of charge.
- The Statistics Bureau will make efforts to improve the web site of the Census results to become user-friendlier by hearing user comments and opinions.

5. Some Issues to be Considered for the Future Censuses

● The use of IT in data collection

- There are many obvious advantages with the Internet response method. One is that those who are familiar with the Internet can submit their responses easily and quickly.
- As the preference of people at large is becoming diverse, providing wider options for response methods other than paper questionnaires is an important consideration.
- The Internet response method is also advantageous for data processing. If the respondents submit their own data by the Internet, there is no need for data capture.

- **Problems with the Internet response method**
 - While the Internet response method is an attractive option for data collection, this option will not be taken in the 2005 Census because of the following problems:
 - (a) How to ensure accuracy;
 - (b) How to achieve cost effectiveness;
 - (c) How to maintain the security and confidentiality of data;
 - (d) How to make the method compatible with the enumerators' work in the field.
 - There are still many problems to resolve before adopting the Internet response method.

● **Consideration of Tailor-made compilation**

- There are various demands for providing more detailed Census tables to exploit the Census data to the extent possible.
- But, it is impossible to compile and provide every cross-classification table because the budget and time for compilation are limited.
- On the other hand, use of micro-data is still restricted in Japan for various reasons. Therefore, by the use of IT, the Statistics Bureau will research a mechanism of tailor-made compilation.

6. Concluding Remarks

- In the Population Census of Japan, new methods utilizing information technology are adopted in various areas except data collection.
- The data collection process is the most important process that determines the quality of data in terms of coverage, accuracy, and timeliness.
- The Internet response method has many difficulties to overcome, and will therefore not be applicable in the 2005 Census of Japan.

- So, the data collection process continues to depend on the conventional method because it is still reliable and affordable in Japan.
- The 2005 Census is to be taken in about seven months from now. The Statistics Bureau will make utmost efforts to conduct the 2005 Census successfully.
- The experiences and lessons learned from the 2005 Census will be reflected in the planning of the 2010 and future Censuses.