18th Population Census Conference
Honolulu, 26-29 August 1998

New Developments in the Dissemination of Census Data in Japan

Yuki Miura
Surugadai University

Background

1 A population census provides comprehensive demographic, social and economic data on population for a country and areas within the country. Many users in different sectors use the census data for different purposes. The Statistics Bureau prepares census data products in various forms and makes them available to users.

2 Population censuses are taken every five years in Japan: full-scale censuses in the years ending with 0 and simplified mid-decade censuses in between. Both full-scale and simplified censuses are complete enumeration and only difference is the number of topics asked. In the 1990 census, principal means of distributing census data were printed census reports and magnetic tapes. For census data not published in the printed reports, anyone can have free access to computer printouts. Some limited tables were offered to users on diskettes. Individual records were not made available to users.

3 So the only means of offering census data to users in a computer readable form was magnetic tapes, with minor exceptions of diskettes. Users who use census data on magnetic tapes have to use mainframe computers and arrange the data sets so as to be suitable for their analytic purposes. This normally requires special computer programming skills.

4 In the advent of an information and communication revolution through a rapid progress in computer and communication technologies in the 1990s, environment of data users has changed. The performance of personal computers has greatly improved and the memory capacity of hard disks has expanded. Newer media such as MO (magneto-optical disk), CD-ROM (compact disk - read only memory), CD-R (compact disk - recordable) have appeared and become widespread, and the internet has become popular media for communication and data distribution. In this advanced environment, users of statistical data are now able to retrieve from appropriate databases the data, which they need and handle those data in their own personal computers using software packages.

5 This paper will discuss the efforts being made by the Statistics Bureau to improve the dissemination of census data products, coping with the changing environment of data users.

Experiences in the 1995 population census

6 The 1995 population census was a mid-decade census, carrying the less number of topics. Tabulation of the results was carried out in several stages. The first data product was preliminary counts of population and households, and this was followed by a 1% sample prompt tabulation. Basic tabulations based on 100% data were carried out in three stages: the first on demographic characteristics of population and households, and housing, the second on labour force status and major groups of industries, and the third on major groups of occupations. Detailed sample tabulations were based on approximately 13% sample data. Data on size and structure of flows of commuting and schooling were tabulated in three stages corresponding to three stages of basic tabulations, based on the place of work or school in relation to place of residence.

7 The amount of data products of the 1995 population census is shown in Table 1. The total volume of numerical data is about 14GB, of which 76 per cent are small area statistics. Tabulation of the 1995 census results has been almost completed by now (August 1998), and the data are made available to users. Except for small area statistics, main results are published in the printed census reports. Statistical tables not included in the printed publications are prepared in computer printouts or on microfiche and offered for public perusal at the Statistics Bureau and the offices of prefecture governments. All tables are also available on magnetic tapes.

8 At the planning stage of the 1995 population census, the Statistics Bureau adopted essentially the same policy for disseminating the census results as in the previous censuses. That is, the media to offer census data to users in a computer readable form are mainly magnetic tapes, and partly diskettes for some data of smaller volume. The only new element was a development and introduction of GIS (geographic information system), that is called CMS (Census Mapping System).

9 However, the Statistics Bureau recently decided to prepare all statistical tables tabulated from the 1995 census results on CD-R. The number of CD-Rs to be prepared is given in Table 2. For data files from each of three basic tabulations and detailed sample tabulation, one CD-R each will be prepared for all Japan and for each of the 47 prefectures. For data files of small area statistics, CD-R will be prepared by prefecture. The Statistics Bureau has developed SI-viewer (Statistical Information Viewer), a browsing and retrieving system of statistical databases on personal computers.

Formats of arranging data files on CD-R

10 It is planned to adopt two types of data file formats: (1) data arranged in database with SI-viewer, and (2) data in EXCEL spreadsheets format. For the 1995 census data, all statistical tables of census results will be arranged in database and recorded on CD-R. In addition, statistical tables derived from detailed sample tabulations will be arranged in spreadsheets and recorded on CD-R in EXCEL format. SI-viewer is similar to the Statistics Bureau's database system SISMAC (Statistical Information System of Management and Coordination Agency) and operates on PCs. It is planned to extend preparation of data on CD-R to the 1980, 1985 and 1990 population censuses.

(Advantages of database format)

11 Arrangement of data in database format on CD-R has several advantages:

(1) Users can easily and quickly retrieve from the database a whole or part of statistical tables they want to use. They can simply follow the menu of SI-viewer interactively.

(2) It is easy to output the data of statistical table image in CSV form (comma separated value). Data in CSV form can be used with popular spreadsheets software packages and therefore can be widely acceptable. Furthermore, data in CSV form are suitable for the purpose of long-term data storage.

(3) Users can operate interactively following the menu on a screen, and display, print or copy the data in a layout of statistical tables. By selecting items and levels of a table head and stub at their own choice, users can design the table, as they like.

(4) SI-viewer can work on PCs with Windows 3.1 or higher version.
(5) Users can have direct access to original data files in database form recorded on CD-R by using popular spreadsheet software, without using SI-viewer. This can be done with PCs either with DOS2.11 or higher version, or with Windows3.1 or higher version, since original data are recorded in DBF (dBaseIIIplus) form.

(Advantages of EXCEL format)
12 Data files are recorded on CD-R in EXCEL97/95/5.0 format. Each statistical table is recorded in one file. CD-R in this format has the following advantages:
(1) Statistical tables are recorded as they appeared in the printed publications and original table forms.
(2) Users can have direct access to data files by using EXCEL5.0 or higher version or any other spreadsheet software that is convertible to EXCEL5.0 or higher.

Data for grid-squares

13 A grid-square is an approximate square of 1km by 1km, delineated by latitude and longitude. In the D.I.D. (densely populated districts), it is further subdivided to squares of 500m by 500m. Data from population censuses as well as from establishment censuses, housing surveys and censuses of commerce are compiled for grid-squares. Topographic maps in scale of 1/25,000 published by Geographical Survey Institute also show grid-squares of 10km by 10km.

14 Data compiled for grid-squares have various advantages.
(1) It is easy to identify the locations of grid-squares on a map, as the grid-squares are defined by latitude and longitude.
(2) Grid-squares are unchanged over time. Therefore, time-series comparisons are possible for data for grid-squares.
(3) Data are also comparable between grid-squares, since surface area of each grid-square is almost the same.
Owing to these advantages, data for grid-squares are widely used by many users both in the public and private sectors for various purposes such as regional development, town planning, measures for prevention of disasters, and improvements of living environment.

15 Data for grid-squares are usually offered to users on magnetic tapes, microfiche, or hard copies. There are about 400,000 grid-squares, of which about 160,000 have inhabitants. Now the capacity of hard disks of PCs is in the order of GB, and therefore data for grid-squares even for a whole country can be handled with Microsoft ACCESS on PC. Users can also analyze the data with the use of GIS (Geographic Information System). Preparation of CD-R for grid-square data is therefore useful to many users.

Use of micro data from the population censuses

16 Discussions are going on in various forums concerning the use of anonymous sample micro data from the censuses and statistical surveys including population census. However, it has not reached the conclusion yet.

Institutional arrangements for dissemination of census data

17 In the present arrangements, the Statistics Bureau distributes printed publications, free of charge, to the agencies of other ministries and local governments. Data files in a computer readable form are offered to the agencies of other ministries and local governments, upon request, free of charge. They can also have access to the Statistics Bureau's statistical database, SISMAC. For users in the private sector and in academic research institutions and universities, both printed publications and census data products in computer readable media are offered at cost through non-profit organizations.

18 At present, a reform of government organizations is in progress. This reform is expected to be implemented from the beginning of the year 2001, and the number of ministries will be reduced from 21 to 12 plus the Cabinet Office by then. It was decided to maintain a decentralized system of statistical organizations, and to accommodate the present Statistics Bureau in the new Ministry of General Affairs to be established by combining three ministries: Management and Coordination Agency, Ministry of Home Affairs and Ministry of Posts and Telecommunications. However, the future status of the Statistics Bureau is not clear yet.

19 Administrative arrangements for disseminating census data will probably be affected by the administrative reform and the resulting organizational changes. Therefore, it is difficult, at this moment, to foresee the arrangements for the 2000 population census data.

Concluding remark

20 Arrangements for dissemination of census data in the 2000 census are not concrete yet, but it is certain that more emphasis will be placed on distribution media suitable for handling on PCs. Further improvements in the dissemination of census data, both in terms of contents and media, are expected, in the light of possible advancements in computer, information and communication technologies in the coming years.

Table 1 Volume of data products in the 1995 population census

	No. of tables	Volume of data (MB)	No. of pages		
			Total	Included in printed publications	Not included in printed publications
1% sample tabulation	50	11.5	1,399	526	873
Basic tabulation (1) (100% data)	124	1,773.7	174,693	32,255	142,438
All Japan			1,391	563	828
Prefectures			173,302	31,692	141,610
Basic tabulation (2) (100% data)	31	553.9	78,685	12,710	65,975
All Japan			520	288	232
Prefectures			78,165	12,422	65,743
Basic tabulation (3) (100% data)	72	427.9	60,791	11,610	49,181
All Japan			944	330	614
Prefectures			59847	11,280	48,567
Detailed sample tabulation (13% data)	27	133.6	16,258	9,672	6,586
All Japan			1,424	504	920
Prefectures			14,834	9,168	5,666
Data on commuting and schooling (1)	27	234.2	29,542	14,374	15,168
All Japan			226	226	0
Prefectures			29,316	14,148	15,168
Data on commuting and schooling (2)	6	128.5	16,213	2,788	13,425
All Japan			27	27	0
Prefectures			16,186	2,761	13,425
Data on commuting and schooling (3)	4	27.1	3,418	3,418	0
All Japan			48	48	0
Prefectures			3,370	3,370	0
Small area statistics	23	10,681.4	1,567,649	0	1,567,649
Basic unit areas	2	1,084.5	106,973	0	106,973
Cho, cho, aza (sections) (1)	13	4,006.9	788,147	0	788,147
Cho, cho, aza (sections) (2)	4	3,423.8	410,852	0	410,852
Cho, cho, aza (sections) (3)	3	2,129.0	255,475	0	255,475
Cho, cho, aza (sections) Data on commuting and schooling	1	37.2	6,202	0	6,202
TOTAL	364	13,972	1,948,648	87,353	1,861,295

Notes:

1 Volume of data relates to only numerical data. There should be additional 78MB required for meta-data and 1MB for computer programmes for data retrieval.

2 The number of pages in printed publications refers to the total number of pages of statistical tables published in the printed census reports. The number of pages of statistical tables not included in the printed publications refers to the number of pages of computer printouts.

Table 2 Number of CD-Rs planned for 1995 census data

	No. of CD-Rs	Remarks
1% sample tabulation	1	
Basic tabulation (1) (100% data)	48	One for all Japan, one each for each of 47 prefectures
Basic tabulation (2) (100% data)	48	One for all Japan, one each for each of 47 prefectures
Basic tabulation (3) (100% data)	48	One for all Japan, one each for each of 47 prefectures
Detailed sample tabulation (13% data)	48	One for all Japan, one each for each of 47 prefectures
Data on commuting and schooling (1)	1	
Data on commuting and schooling (2)	1	
Data on commuting and schooling (3)	1	
Small area statistics	189	
Basic unit areas	47	One each of 47 prefectures
Cho, cho, aza (sections) (1)	47	One each of 47 prefectures
Cho, cho, aza (sections) (2)	47	One each of 47 prefectures
Cho, cho, aza (sections) (3)	47	One each of 47 prefectures
Cho, cho, aza (sections) Data on commuting and schooling	1	
TOTAL	364	