

Singapore's Approach to Census 2000

Singapore Department of Statistics

I. Introduction

1. The United Nations (UN) recommends that a national census be taken at least every 10 years. As the value of census data is increased if it can be compared internationally, the UN further adds "countries may wish to undertake a census in years ending in '0' or as near to those years as possible".
2. Singapore's first census was taken in April 1871 as part of the Straits Settlement Census. Since then, regular censuses were undertaken at ten-year intervals up to 1931. The Second World War delayed the next censuses till 1947 and 1957. Singapore's first population census after independence was conducted in 1970 in line with United Nations' recommendations to designate years ending in "0" as census years. The next two censuses were conducted in 1980 and 1990.

II. The 1970 and 1980 Census

3. The 1970 and 1980 Censuses followed a traditional fieldwork approach. In the first stage, houses were numbered to ensure complete coverage. The second stage involved a large number of field interviewers visiting the households to collect the information with paper forms and pens. The large volume of information collected was then processed through a cycle of coding, data entry, verification and table generation.

III. The 1990 Census

4. In the mid-1980s, the People Hub database was set up with unique identification number (UIN) for each Singapore citizen and resident.
5. The 1990 Census capitalized on the potential of the UIN for record linking and made use of the People Hub as the basis of conducting the census. Information captured in People Hub was merged with a few other administrative databases. As far as possible, the census forms were pre-printed with data from the databases for verification with respondents.

6. Field interview was still the main method of data collection. However, respondents were allowed to self-enumerate (i.e. fill up the forms themselves) and return the forms to the interviewers. About one-third of the respondents chose self-enumeration.
7. As in previous censuses, the data collected were processed through a cycle of coding, data entry, verification and table generation. The 1990 Census exploited the then database technology running on a Fujitsu mainframe for data processing. Coding was done by batch mode and extensive data verification rules were drawn up for batch mode checks on census records. Table Producing Language (TPL) was used to generate the census tabulations.

IV. Census 2000 – A register-based census

8. Since 1990 Census was the first time that a database was used to conduct the Census, it was deemed necessary to verify the information with those collected from the field. The results were encouraging. Following this, a Household Registration Database (HRD) was set up. Information in the HRD originated from the 1990 Census and is updated regularly by administrative data from various sources.
9. In many countries, a population census is conducted together with a housing census to find out the characteristics of dwelling units. Since 1980, DOS maintains an up-to-date database on dwellings. In 1996, this database has been upgraded and renamed National Database on Dwellings (NDD). The NDD and HRD together give a physical location for every household in Singapore.
10. With the basic or core items on individuals and houses being available from the HRD and NDD, it would suffice to conduct a register-based census in the year 2000. Additional data required for in-depth studies will be collected from a large sample of the population. Experience from the past censuses and sample surveys indicates that a 20% sample would provide sufficient details for in-depth studies and meet the need of the majority of users.

V. Coverage

11. A traditional census enumerates all persons within the territory or country at the designated reference time known as the “census day”. This is known as the “de facto” census. The “de facto” census has the advantage of easy implementation. The exclusion of residents who were temporarily overseas and the inclusion of foreign visitors do not pose a major problem with limited international travelling in the past.
12. A “de jure” census on the other hand, enumerates all persons at their “usual place of residence” at the designated reference time. This will theoretically cover all residents who are temporarily out of the country. “Temporary or transitory” visitors and “non-locally domiciled persons” are excluded. However, it may be difficult to define “usual place of residence” as well as “temporary or transitory” visitors.
13. A strictly register-based approach to Census 2000 meant that the population count would in fact be “de jure”. All Singapore residents who are overseas will be included in the total population count as their records will be in the database. Similarly, foreigners living in Singapore will be included, as their records will be merged into the database from administrative sources. “Temporary or transitory” visitors and “non-locally domiciled persons” will be excluded from the total population count. However, following past census practice, a special count will still be conducted for these groups for record.

VI. The Tri-Modal Data Collection Strategy

14. For the 20% sample enumeration, Census 2000 will adopt a tri-modal data collection strategy comprising Internet enumeration, Computer Assisted Telephone Interviewing (CATI) and fieldwork (with mail-back option).

Internet Enumeration

15. The option of Internet Enumeration will be made available to all households selected for the 20 per cent sample of the population.

16. Upon the launch of the publicity campaign, all selected households will receive a notification letter with a password. Using the password and the UIN, respondents who wish to be enumerated by Internet would be able to log-on to their household record in the database via the Census website (www.census.gov.sg).
17. Some basic data already in the pre-census database will be displayed. The respondent would then proceed to fill up the rest of the census questionnaire on-line. Various user-friendly help features and explanatory notes would be provided. The system will also perform simple on-line checks, and prompt the respondent to re-enter data that are clearly wrong or inconsistent.
18. Respondents will be given the option to save and exit from a partially completed questionnaire and fill up the remaining questionnaire at a later time. Security features will be built-in to prevent unauthorized access, hacking or jamming of records over the Internet.

CATI

19. CATI was first deployed in the 1995 mid-decade General Household Survey (GHS) covering about 300,000 Singapore residents. In the survey, some two-third of the households were successfully interviewed by CATI. The Department intends to build upon the success of the 1995 GHS, and exploit CATI as the main mode of data collection in Census 2000.
20. The CATI system allows the interviewer to perform multiple tasks of interviewing, data entry and simple coding simultaneously. With most questions in multiple-choice format, the interviewer needs only to point and click on the right answer. The interactive environment also allows for automatic branching of questions. For e.g., should the respondent be a full-time student, the system will skip questions on economic activity and move on to ask questions on transport mode to school. Answers are automatically coded wherever possible and updated into the database.
21. Households that have not submitted their returns by Internet will automatically be scheduled and dialed up for CATI interview after a cut

off date. Households with unlisted or without telephone numbers could still opt to be enumerated by CATI by calling the Census Hotline.

22. Like the Internet enumeration system, various help features and explanatory notes will aid CATI operators. The CATI system will incorporate streamlined questioning, It will also feature on-line checks, and prompt the operator to re-enter data that are clearly wrong or inconsistent.

Fieldwork

23. Records will be scheduled for fieldwork if it could not be contacted by CATI after a fixed number of telephone attempts. These records will be grouped by areas and passed to regional census offices.
24. Fieldworkers will visit these remaining households to conduct face-to-face interviews. Should they fail to contact these households, they would leave blank census forms with these households who could fill and mail the form back to Census Office. All forms coming back from the field will be imaged and the data captured through OMR software. Data entry will only be necessary for descriptive fields.

Cost of the various modes

25. The cost to Census Office varies according to the mode of response chosen by respondent. Field enumeration cost the most, as it is the most labour intensive. This is followed by the mailing method, which requires census officers to scan in the forms returned and perform corrective data entry for descriptive items. The expected high proportion of incomplete forms also meant that census officers had to contact the household to fill up the missing data items. CATI interviews are cheaper as data entry is directly done during the interview and no transport time or costs are incurred. Internet self-enumeration is the most cost efficient, given a properly designed system, as the respondents perform data entry.
26. It is difficult to estimate the proportion of households that will be enumerated by the various methods. The Internet penetration level in 1996 was only 8.6% of total households¹. Furthermore, Internet

enumeration requires respondents to take the initiative and play a proactive role.

27. On the other hand, a higher proportion of households would already have Internet access in their workplace or school. The PC penetration rate stood at 36%¹ in 1996. With the trend of PC vendors combining PC sales with modems and Internet access, it is likely that Internet penetration will move up towards PC penetration levels. The many national initiatives taken by the Government to provide Singapore with an island-wide state-of-the-art information infrastructure will also increase the number of households with Internet access significantly.
28. Furthermore, the Inland Revenue Authority of Singapore (IRAS) has recently launched E-Filing as a means by which tax payers could submit their income tax declaration form via the Internet. The response from tax payers have been encouraging. With the trend of government agencies moving in the direction of electronic transactions, it is likely that the population would become increasingly receptive to the idea of providing census information via the Internet.

VIII. Data Processing and Verification

29. Once data have been collected and stored, the coding of descriptive items such as occupation and industry begins. At the same time, data verification is necessary to rectify inconsistencies, duplicates or errors in the records. The experiences of GHS95 and Census 1990 showed this to be very time consuming. Officers verifying the records often have to re-contact respondents to sort out inconsistencies in the records.
30. Census 2000 will exploit the increasing computing power of the PC by having enhanced data verification checks at the front end. The CATI and Internet systems will have more extensive on-line checks for inconsistency and prompt the interviewers to correct any data entry errors on the spot. By shifting more verification checks to the front end, more errors could be corrected at the point of data collection, where the opportunity to double check with the respondent is available, rather than at the backend where re-contact with the respondent is costly.

¹ IT Household Survey Report 1996 by NCB

31. To handle the coding of descriptive items, the Department had tied up with Kent Ridge Digital Labs (KRDL), a research institute, to develop the Advanced Coding Environment (ACE). ACE comprises two distinct areas, namely the auto-coder and the coding wizard software. The auto-coder performs a direct string match with a dictionary of codes. All records with distinct and non-ambiguous job titles would be automatically coded in this way.
32. Records that could not find a perfect match in the auto-coding phase would be channeled to computer-assisted coding. At this phase, the coding wizard provides intelligent assistance to the human coders in searching for the correct codes. Besides performing sophisticated string match, the coding wizard engine would take into account related fields such as the highest qualification, income and age group in determining what are the most likely codes for industry and occupations. The wizard then list out the most likely codes, in descending order of likelihood. The coder need only analyse the record and pick the correct code.

IX. Data Storage and Analysis

33. Data warehousing concept will be used to manage the vast quantity of data efficiently. Database warehousing is a major driver in IT presently and offers a data storage architecture for collating, processing and managing data from different sources into a single repository so that analysis can be performed.
34. In addition to basic statistical tabulations, data mining tools will be used during the analysis stage to maximize the usage of the vast amount of data. With the rapid changes in IT technology, it will be prudent to keep abreast of the latest development in new tools and programs and to finalize the strategies nearer the data processing stage.

XI. Manpower and Training

35. The register-based approach to Census 2000, together with the innovative use of the various technologies meant that only one-sixth of the total

number of census workers deployed at the height of the 1990 Census would be required for Census 2000.

36. However, with multi data collection modes, it is necessary to recruit census officers with better educational profile and IT skills. Training of census officers will be given to equip them with the skills to handle the various computer systems and software. In addition, census officers have to be trained on concepts and definitions, line of questioning, responses to respondents' queries and soliciting complete and reliable answers.

XIII. Discussion and Conclusion

37. The register-based approach to Census 2000, supplemented by a large-scale survey as described in this paper, will mark a watershed in the history of census taking in Singapore. For the first time since 1871, information will no longer be "canvassed" from the entire population. This is in line with the approach taken by Denmark, Finland, Norway, Sweden and the Netherlands.
38. Outside of Europe, Singapore would be the first country to embark on the register-based approach. In deciding to move in this direction, the Department of Statistics had studied three key issues. First, the quality of administrative data in Singapore is sufficiently high to produce an accurate count of the population and its basic characteristics. Secondly, the legal framework and data confidentiality practices in Singapore permit the sharing of various administrative information. Finally, the cost savings in adopting this approach are substantial. It is estimated that the cost of conducting a register-based census, coupled with a large-scale survey, is only 40 per cent of the cost of a full scale census.
39. Of the 163 censuses taken in 1990 round, only 23 countries used more than one method of data collection. Of these, only 2 countries adopted a combination of three data collection methods. The tri-modal data collection strategy adopted for the 20 per cent sample enumeration is therefore a bold experiment in multi-mode data capture and the application of cutting-edge technology. Department of Statistics views the integration of the various modes as a critical success factor for Census 2000. To ensure smooth workflow and seamless transfer of data from one

mode to another, a census management system will be built to track and move records from one phase to another.

40. The heavy investment in IT for Census 2000 is expected to bring significant returns in the future. The integrated use of the various technologies in Census 2000 will set the foundation for the Department's IT Vision for the future. This vision seeks to provide a holistic solution to the entire workflow in data collection, processing and publication. Through an integrated electronic system, data could be collected, processed and tabulated seamlessly, so that the average turnaround time in the delivery of output to the users will be vastly reduced.
41. Beyond 2000, the Department of Statistics will look into a system of continuous measurement of the population by tapping on the records of the HRD and the NDD. A system of regular small-scale surveys will be put in place to collect information not obtainable from administrative sources and to monitor population and social trends of current interest.

Department of Statistics
Singapore

References:

1. Emerging Issues Related to the 2000 World Population and Housing Census Programme, by Sam Suharto (UNSD, Technical Notes, Dec 96)
2. Use of Administrative Records in Population Censuses and in other Demographic and Social Statistics, by Sirageldin H Suliman (UNSD, Technical Notes, Nov 95)
3. Handbook of Population and Housing Censuses, United Nations
4. The Register-based system of Demographic and Social Statistics in Denmark – an overview, by Lars Thygesen (Statistics Journal of the United Nations ECE 12 1995)
5. The use of Identification Numbers to link Information from Various Sources and Create Alternative Statistical Units and Concepts, by Finn Spieker (Statistics Journal of the United Nations ECE 12 1995)
6. Evaluation of the results of the Register-based Population and Housing Census 1990 in Finland, by Riitta Harala (Statistics Journal of the United Nations ECE 12 1995)
7. Which Countries will Follow the Scandinavian Lead in Taking a Register-Based Census of Population?, by Philip Redfern (Journal of Official Statistics, Vol 2, No 4, 1986)