

SOME IMPROVEMENTS TO 2011 POPULATION AND HOUSING CENSUS IN SRI LANKA

**H.R. GUNASEKERA
DIRECTOR
DEPARTMENT OF CENSUS STATISTICS – SRI LANKA**

**Paper Presented to the 25th Population Census Conference
Seoul, Korea
May 25 – 27, 2011**

Some Improvements to 2011 Population and Housing Census in Sri Lanka

1. Introduction

Sri Lanka has a long history of Census taking. The first of a scientific series of decennial Census was conducted in 1871. Five regular Censuses were held thereafter up to 1921, establishing a well defined Census methodology. The census of 1931 was restricted to a head count due to financial reasons arising from economic depression. The Census due in 1941 was postponed until 1946 on account of World War II. A paper shortage resulting from the Korean War caused a rescheduling of the Census from 1951 to 1953. The census in 1960's was held in 1963. The practice of conducting Census in years ending with digit 1 was reestablished in 1971, which marked the 100 years of experience in conducting Censuses. The twelfth of the series was taken in March 1981, but the Census due in 1991 could not be held due to the disturbed conditions prevailing in the Northern and Eastern provinces of the island. The last Census in 2001 was conducted after a lapse of 20 years from previous Census, which is the largest time gap recorded for an intercensal period in the history of Censuses in Sri Lanka. However, Census 2001 was able to carry out completely in 18 out of 25 districts due to the disturbed conditions in Northern and Eastern provinces of Sri Lanka. The peace has drawn to Sri Lanka in 2009. The Population and Housing Census to be carried out in 2011 is of special significance as it is able to cover the whole country including Northern and Eastern Provinces after 30 years.

2. Conducting the Census

Census 2011 is carried out in 3 stages

1. Mapping
2. Listing all buildings
3. Enumeration of Population and Housing

To avoid omissions or duplications of buildings units, maps were drawn at smallest administration division level (Grama Niladhari Division) with clear identifiable boundaries. Such divisions were further divided into Census blocks. A Census block comprised of around 150 building units. The Census block boundaries were also marked on the map during the mapping operation. Based on the maps prepared, all the building units were listed in a separate form during the listing operation. All building units were classified into housing units, collective living quarters, and non-housing units and listed in convenient and suitable order.

During the enumeration, Census enumerators collect information pertaining to every individual who usually live in the household. Housing Information is also to be recorded.

3. Improvements to the Census 2011

Based on the lessons learnt from the previous Censuses, certain issues and problems in data collection methodology, processing and dissemination were identified. Such concerns were addressed to the extent possible using our own experience and taking into account the international best practices.

3.1 Enumeration Procedure

In past Censuses enumeration was done in two stages i.e Preliminary Census and Final Census. During the preliminary Census, information was collected from each household based on usual residence approach. Final Census was conducted between 6 p.m and 12 midnight after about 1 week from preliminary Census. During the final Census, information collected in the preliminary Census was updated by identifying those who were not present on that night and including the information of visitors present. In addition, special 'outdoor enumeration' was carried out for those who were on move that night. The data tabulated were based on the results of final Census. So essentially, past Censuses in Sri Lanka were conducted on 'defacto' basis. i.e persons were enumerated wherever they happened to be at the place of final enumeration irrespective of their usual residence.

Conducting The Census on defacto basis as described above created many issues. Conducting the final Census within 6 hours created many operational problems. During this short period of time large number of enumerators (about 110,000 in 2001) had to be mobilized and organized to visit all housing units. Further, additional outdoor enumerators (about 8000 in 2001) were appointed to cover the persons who live on the streets (homeless) and those who happened to be travelling and remain at buses, trains, bus stands, railway stations, busy places etc. Government requests to the public to stay at home to make the outdoor enumeration within manageable level. Still in the current socio-economic set up it is very difficult to keep public at home. Appointment of outdoor enumerators incurred an additional cost as well.

On the other hand, many data users need the population information on the basis of usual residence rather than on place where present. This is particularly useful in the policy planning purposes in the areas of education, health etc.

Several studies were conducted and advices from international consultants were obtained to study the problem and find a solution. Finally to overcome the above difficulties, it was

decided to base the enumeration on de jure concept. But in order to make some comparative analysis with the last Census it was further decided to adopt 'recall de facto' method to find out whether there are any usual residents absent or temporary visitors present at the reference time of the Census. The key here is to ensure proper recall; therefore the sooner the enumerators visit the households after reference time, the better the memory of respondents. This will avoid the mobilization of large number of enumerators in a very short time, and no outdoor enumeration is needed. However, people without usual residence (homeless persons) should be enumerated at reference time.

So the new enumeration phase consists of two stages i.e Enumeration round and Revision round.

Enumeration round: Information is collected from all usual residents in units where people reside, 3 weeks prior to the Census. In addition, housing information is also collected.

Census moment refers to 12 midnight of the day immediately before the Census day (or 00.00 hrs of Census day). Persons who do not have a usual residence (homeless) are enumerated at the Census moment. Taken into account the practical difficulties this is to be done few hours before the Census moment.

Revision Round: Revision is done on two days, i.e on the Census day and the following day. All housing units and collective living quarters covered during the previous stage are visited. In the revision round

- Usual residents absent and temporary visitors present at Census moment are identified and record the relevant code in the Census schedule. Information is recorded for the temporary visitors. This will enable to get the de facto count of the population in the enumerated units.
- Information collected is revised to account for any usual resident who had died or any new birth in the household that recorded between Census enumeration and the Census moment.

3.2 Census Data Capture

The current system of data capture is centralized key-to-disk procedure. This causes considerable delays in data processing. In addition to delays, data entry can be a potential source of error unless carefully planned verification procedures are adopted. Application of image based form processing technology such as OMR/ OCR/ICR can be used to expedite the release of the census data. Before implementation of any such technology greater attention should be given to examine its suitability. The success depends on the extent to which the

necessary conditions prevail such as the quality of the specially printed questionnaires, training given to enumerators in recording the responses clearly and of course on the selection of technology considering the uncertainties. Such procedures need substantial allocation of resources in terms of equipment and manpower.

Census scan processing system (CSPS) is outsourced as it is felt that the equipments become outdated soon as the technology is changing at rapid pace. A company has been selected through a lengthy procurement procedure. Census schedule has been formatted to suit the scanning process and the printing of schedules is to be given to the same company.

Whenever OMR/OCR/ICR engines do not yield the expected results in the conversion process the system has the ability to use alternative process. CSPS will output the data in ASCII and CPro formats with a unique identifier. The modules work on a Local Area Network based on TCP/IP protocol. The system includes an automotive and highly reliable backup and recovery subsystem to be deployed in adverse situations. CSPS has number of verification mechanisms. To evaluate the performance of the scanning process sample data files and images are checked 5 percent in the first fortnight, 2 percent in the second fortnight and 1 percent thereafter. It is expected an accuracy rate of 99 percent for OMR and 95 percent for OCR and ICR. The batches with less than that accuracy will be reprocessed. A pilot test is carried out covering about 10,000 housing units and approximately 50,000 persons in order to thoroughly examine the scanning process.

3.3 Computer Assisted Coding

The usefulness, quality and timelines of the Census outputs depend on the methods adopted to process the data. Although most of the items of the Census questionnaire are pre-coded, certain items need to be coded before computer processing begins. Occupation and industry are two items usually need to be coded. Past experience reveals that this is laborious, time consuming and can be an important source of error. Software has been developed to assist the coders in finding the most suitable code for particular occupation or industry classification. This system has been tested very successfully using pilot Census data. This computer assisted coding system (CAC) will greatly reduce the time taken for coding and the risk of having errors.

3.4 Computer Editing and Imputation

The usefulness, quality and timeliness of the census outputs depend on the methods adopted to process the data. Editing and imputation is another area that is extremely important to achieve the above characteristics. The practice in the last census was to get error print-outs followed by manual corrections and re-entering the correct entries. This process is very

laborious and time consuming. To expedite this operation online editing and more carefully tested imputation procedures are being used under expert guidance.

3.5 Master Register of Census Blocks

Master Register of Census Blocks (MRCB) is an automated data base written in ACCESS consisting of the information at census blocks with all the geographic details and a unique identification number. This has been prepared at the time of demarcating the census blocks. The basic information such as number of housing units, population etc collected at mapping and listing stage can be entered into the data base as and when such data are available. MRCB is an invaluable tool for monitoring the progress of field work at different stages of the census, planning for material and human resources needed at Census block level, as well as getting various summary reports for dissemination purposes. As the data entry for MRCB is decentralized at district level various security measures have been designed to safeguard the integrity of data.

3.6 Dissemination of Census Information

Improving the census dissemination system is another challenging task. Currently census information is disseminated through printed as well as electronic media and also via internet. But information in the internet is limited to predefined set of static tabulations. This will be continued in future for information which is in heavy demand. However there is a greater demand for user defined interactive tabulations together with thematic mapping. To achieve this objective Census-info package has been tested with test data using the 2001 Census.

3.7 New Definition of Urban Areas

Urban areas in Sri Lanka are defined on the basis of administrative boundaries. During early 1980's urban areas comprised of all Municipal Councils, (MC) urban Councils (UC) and Town Councils (TC). In 1987, Town Councils were abolished and absorbed into Pradesheeya Sabahs, which were essentially rural areas. Since then only MC and UC areas have been considered as urban areas. The Population and Housing Census conducted in 2001 also used this definition. As a result, the urbanization in Sri Lanka is underestimated and does not actually reflect the ground situation. There are areas outside the administrative boundaries of MCC and UCC with highly urban characteristics. Hence it is very important to have a new definition of urban areas considering the current socio- economic set up.

During the stage of listing of buildings of the Census, which is just completed, a questionnaire was canvassed to collect the basic information pertaining to Grama Niladhari (GN) Divisions. It collected information on basic facilities available such as whether there is school, hospital,

dispensary, post office, common markets, bank, etc and number of such facilities in the GN division. This information can be used to develop a statistical definition of urban areas using multivariate analysis of variables related to urbanization. The final conclusion on the definition has to be taken as a joint effort of all stakeholders dealing with urban development and planning.

3.8 E- Census

It is planned to carry out an E-Census on a small scale during the 2011 Census. The current practice in collecting the information is to visit each and every household by enumerators. It is a known fact that certain segments of population, particularly most affluent people living in Colombo city do not co-operate well with the Census enumerators. E-Census is targeted to reach a sample of such population. The consent in the participation is taken at the house listing operation of the Census. User ID and a password is given to household who have given consent to participate in the E- census. They can download E-Questionnaire (E-Q) from the web and submit it to the web server which contains the E-census data base. The confidentiality of data in E-Q'S are safe guarded through the unique ID and password provided.