



**United Nations International Seminar
on Population and Housing Censuses:
Beyond the 2010 Round
27-29 November 2012
Seoul, Republic of Korea**

SESSION 7: Use of modern technologies for censuses

**USE OF TECHNOLOGY FOR CENSUS DATA ANALYSIS AND DISSEMINATION WITH
SPECIAL REFERENCE TO THE CENSUSINFO DASHBOARD**

Prepared by

**Ms. Hekali Vikiye Zhimomi
Government of India**

* This document is being reproduced without formal editing.

**Use of Technology for Census Data Analysis and dissemination with special reference to the
Censusinfo Dashboard**

Introduction:

The Indian census has a rich tradition and enjoys the reputation of being one of the best as well as the biggest exercise in the world. The first and complete census in the country was conducted in the year 1881. Since then, a nationwide census has been conducted every decade. India, which is home to more than a sixth of the world's total population with a figure of 1.2 billion, completed Census 2011 which is its 15th census. Census in India is carried out in a traditional manner which involves a house to house survey conducted by enumerators who canvas questions and records the information furnished by each household in a questionnaire.

Census 2011 covered 35 States/Union Territories, 640 Districts, 5,924 Sub-Districts, 7,935 Towns/Cities, 6,40,867 Villages (.64 million villages), 240 million households and 2.3 Million (23.56 lakhs) EBs. It is carried out in 2 phases- the Houselisting and the Population Enumeration. During the House listing Phase, a questionnaire (also called Schedule in census terminology), 35 questions were posed to the usual residents. Questions asked include details of the house – its use, building materials, assets and amenities in order to assess the condition of human settlements, housing deficits, etc. The population enumeration which followed the Houselisting phase consisted of 29 questions about the residents' personal details, education & economic status, etc. After the population enumeration phase, a revision round is conducted, during which any changes in the entries that arise on account of births, deaths and migration between the time of the enumerators visit and a reference date/time is noted down and records updated.

The Census 2011 in India involved approximately 340 million census schedules (questionnaires) in 16 different languages, 6 million instruction manuals in 18 languages consuming 8000 MTs of paper despatched to about 17000 locations . The cost involved in carrying out this mammoth exercise was just a tenth of the world average at less than \$0.5 per person as compared to the estimated \$4.6 per person according to the 25th Population Conference organized by the UN statistical Division held at Seoul in 2011. The exercise spanned over a period of 11 months involving 2.7 million field functionaries (Enumerators and Supervisors).

Data Processing & Analysis:

On completion of the census taking process, the important tasks of processing, analyzing and dissemination of the data collected followed. In Data processing, India has always been in the forefront of using the latest technologies. For Census 2001, India was one of the first countries to use the image based Automatic Form Processing Technology taking the help of High Speed Duplex Scanners for image capturing. In Census **2011**, India has adopted more developed ICR Technology with advanced features. In particular, India is using the unique TILE module which optimizes data accuracy with a systemized display of characters grouped together to allow easy identification. This system makes it possible to identify which characters are correct and which are not and allows the marking of rejected characters. It also ensures more accuracy in data processing.

Data Analysis in the Indian census operation has always been a mammoth task given the volume of data involved. With data collected from across 35 States, consisting of 640 Districts, 5,924 Sub-districts, 7,935 towns and 0.64 million villages, the ORGI has the challenging task of systematically analysing and presenting the data. As has been the tradition, the ORGI creates two separate databases, the House listing & Housing Census Data (at Household level) and the Population Enumeration Data (at individual member of the Household level). In Census 2011, an attempt is being made to link these two databases to cross-tabulate information e.g cross tabs on Condition of Housing with Economic Condition, etc. To enable this, the boundary of the Enumeration Areas (EA) have been kept unchanged during the two phases of the Census operation and provision has been made in the Household Questionnaire (Phase 2 Operation) to record the Household Number marked in the Phase 1 Operation. This ensures that the Enumeration Area and the Household Numbers serve as link fields in the two databases. This important link will enable us to have micro information that will help planners to evolve need specific programs and enable better identification of potential beneficiaries/ target areas. As boundaries of Enumeration Areas (EAs) are not permanent it is not possible to link the EA from one census to the next as a result of which every census has generated stand-alone databases. Although the ORGI has made an effort to generate and release time series tables from previous census there are challenging issues as many jurisdictional changes occur, new villages and towns are created. Hence an attempt is underway to link the databases available since 1991 Census on jurisdictional changes up to Town and Village levels in the country.

Following the completion of the Census 2011, Provisional Population Totals for India and States have been compiled from the Enumerator's Abstract manually and declared within four weeks. Scanning

of filled-in schedules which have been collected back from the field was done in two phases - – House listing & Housing Census and Population Enumeration, and has been completed. Extensive quality check and data validation has been undertaken at each stage of data processing and analyses, using the CS Pro Software.

In addition, within a year of completion of the census field operation, The Office of the Registrar General of India has already published Tables on Houses, Household Amenities and Assets. The task of data checking and validation has been made far easier and quicker with the aid of technology and what normally would have taken more than several months can now be completed in a month's time, making the data far more relevant to current usage.

Data dissemination

The Government of India relies heavily on the Census data in formulating its Five-Year plans, Annual plans and various other welfare schemes. Questions canvassed during the Census covers a wide gamut of information pertaining to every resident. Data drawn through this exercise forms an important basis for formulating socio-economic and administrative policies. It is the biggest source of primary data on: Demography, Economic Activity, Literacy & Education, Housing & Household Amenities, Urbanization, Fertility and Mortality, Scheduled Castes and Scheduled Tribes, Language, Religion & Migration. The user profile of census data includes policy planners, researchers/academicians, voluntary organizations, and businesses/industries right down to the grass root level.

To ensure that the vast amount of data collected during Census is made optimal use of by potential users, it is imperative that data is made easily accessible in various modes/formats to cater to a wider section of users. To meet this objective, and to cater to the ever growing number of data users, in 2002, a Data Dissemination unit (DDU) was set up in all the Directorates across the country to disseminate census data to different types of users including administrators, policy planners, voluntary organisation, commercial organisations, researchers, etc. Following 2001 Indian Census, the Office of the Registrar General India framed a Data Dissemination Strategy to be followed in all the Directorates to disseminate data more efficiently. Consequently all the 33 Directorates have set up a Data Dissemination Unit equipped with the best and latest facilities.

The modes of Data Dissemination through different Data products include:

- Printed reports (tables, State profiles, Analytical reports, Data Sheets on important topics, Atlas, Booklets)
- CDs, DVDs including subject specific CDs
- Website and downloadable files in the official Census website. An important part of the Census website relates to GIS application which extends the facility of generating thematic maps based on 2001 Census data.
- Social networking site – Facebook/Twitter
- Workshops/seminars to sensitize data users about availability and use of census data. Series of Data Dissemination workshops and seminars are being held every year to promote the usage of Census data to different types of users and sensitize potential users about the availability of a wide array of demographic, economic and education data.
- An intranet facility has been established within the organisation for faster flow of Census data and other information across the country. This serves the data users located in different parts of the country and outside.
- Following the completion on Census 2001, data was disseminated using CensusInfo through an interactive CD containing data on Demography, Economy and Education, Time Series Data (from 1991 to 2001) in the form of tables, Graphs and Maps. The Provisional data of Census 2011 has already been made available on CensusInfo India Dashboard at the Census of India website. Data on Census 2011 including Data on Housing and Household Amenities and Assets (data which were collected during House listing Phase) and Primary Census Abstract (which is a publication containing data pertaining to each District in the country) will be disseminated using CensusInfo India through interactive CD and the Census of India website.
- As part of the 12th Five –year Plan, the Office of the RGI, with the support of UNFPA, is implementing a scheme in 2012 to set up workstations in the ORGI as well as in all the 33 Directorates of Census Operations (in 2 phases) for the purpose of setting up of a Digital Archive of Old Census Reports (1872 to 2011) for use by researchers and other users. These workstations will be used to store census micro-data and serve as a hub for Data Dissemination. This exercise entails digitization of existing Census records in the form of printed reports, microfiches, microfilms, etc by scanning each page and then converting them into PDF. To access this huge collection of digitized

census data, retrieval software would be used. Once the digital archive is ready, facility will be in place for researchers and various other users to access the archived reports online as well as offline.

- A software module on Decision Support System (DSS) is being developed which will provide access to latest census data at District, Block and Village/Town levels for planning purposes. Datasets to be used for DSS are the Primary Census Abstract at village/town level and Village & Town Directory data collected through State Governments and would involve linking of village level data with maps. This will greatly help planners in identifying less developed regions/villages in terms of infrastructure and amenities.
- The most effective way to promote the use of census data is to introduce the idea to future users, i.e, students. Keeping this objective in view, it has been proposed that School Kits on Census Data containing reading materials would be prepared in attractive designs and styles in different languages and distributed to at least 100 schools per district in all the 640 districts in the country. It is envisaged that this exercise will go a long way in promoting the importance and usage of readily available raw data across different industries and segments.
- A workstation for researchers/academicians has already been set up at Jawaharlal Nehru University. 18 more such workstations are being set up in collaboration with different Universities in different parts of the Country. The objective is to allow researchers to undertake research on database of more than a billion persons as collected in Census after anonymizing the sensitive fields.

Unicef has been supporting Data dissemination activities of Census of India 2011 results which includes development of dissemination strategy and software support for dissemination of census data in user friendly manner to enable quicker and easy dissemination of census data to users and promote use of census data.

CensusInfo India and its advantages

For the first time in 2002, the Office of the Registrar General, India, has used the CensusInfo to disseminate census data. CensusInfo is a database system, developed by the United Nations Statistics Division, in partnership with UNICEF and UNFPA. Censusinfo has been adapted from the DevInfo database technology and distributed free to countries for presenting and dissemination of Census Data. “The primary objective of CensusInfo is to improve the dissemination and utilization of

population and housing census results at any geographical level with the mapping and graphs facilities. Users can easily retrieve required data from the database in the form of tables, graphs and maps from across different indicators and customized disaggregation.” Every Census Directorate in the Country was asked to hold Workshops to popularize CensusInfo and also to sensitize users about the usage and availability of Census 2001 data. Representatives from Government, NGOs, Universities, Public and Private Enterprises participated in the workshops and wide awareness was created to enable the usage of the information [provided through Censusinfo at various levels. CensusInfo was further updated with Data on Housing amenities and Assets – 2001 and Primary Census Abstract – Census 2001.

Following the Census 2011, The Office of the Registrar General of India has already launched the Census Info Dashboard at Census of India website to generate quick profiles on Provisional Results released from Census 2011. The Provisional data of Census 2011 has already been made available on Censusinfo India dashboard at the Census of India website. Users can access State data with a single click of the button, and by double clicking on the particular State, the District data on various indicators such as total population size, child population, population density, sex ratio, literates and literacy rates. Users can also access disaggregated data on the basis urban and rural areas across different levels of geographic locations. Shortly, data on Housing and Household amenities and assets (data which were collected during the House listing Phase) and Primary Census Abstract (which is a publication containing data pertaining to each District in the Country) will also be disseminated using CensusInfo India.

India being a multi-ethnic (having more than 2000 ethnic groups), multi-lingual, multi-cultural country with every major religion being represented, CensusInfo has been a boon in accessing, retrieving, storing and dissemination of data efficiently and effectively.

Some advantages of CensusInfo, particularly for India, are:

- **User friendly features:** One of the biggest advantages of this software is that a user possessing any level of computer proficiency can access/retrieve information on various indicators presented in high-quality tables, maps and graphs.
- **Easily accessible:** Users can quickly access data (in print, CD and Web version) on the basis of age/sex, Rural/Urban, and various other indicators including different time-periods for comparison purposes. Users can easily generate tables, graphs and maps. The software uses aggregated data, while easily handling data disaggregated by geographical area and various sub groups. It also supports exchange of data with other software applications

- **Flexible software:** The flexibility of the software to adopt country specific application is another huge advantage. India being a multi lingual country, the module which provides support for local languages will be of great advantage to users.
- **Supports unlimited geographical levels:** CensusInfo supports unlimited number of geographical coverage down to village level, which is another very useful feature considering that close to 70% of Census Data pertains to village level.
- **Adding new indicators & customization of Meta data:** The feature which allows for countries to set up their own indicators with the option of adding new indicators and customizing the metadata for each indicator is another advantage. Indicators can also be expressed in different ways and expressed as counts, percentage, ratios, proportion, rates or averages.

To popularize the usage of the Censusinfo Dashboard, a National workshop has been planned at Delhi for 26th and 27th November, 2012 during which CensusInfo 2012 will be launched by the Union Home Minister, India. During the workshop, training on CensusInfo will be conducted where demonstration of the software as well as hands on training on generating tables and graphs, using map for exploratory analysis and preparing presentations will be imparted to the participants. The ORGI has already planned to organize workshops for prospective users in every Directorate in respective States within 15th December, 2012.

Conclusion:

As the publisher of Census Data in India, the Office of the Registrar General, India, with the support of various agencies of the UN, has been constantly upgrading itself to disseminate data effectively to various types of users. With the help of CensusInfo 2012, it is hoped that the Census data will be made more easily accessible and more users will be able to take advantage of the free data for various purposes. However, for data dissemination to be made truly accessible to all, it is important thing to bear in mind that there are certain sections of society in any country who would not have access to technology, no matter how user friendly. For a country like India, where almost 70% of its population reside in the rural areas, it is a matter of concern and challenge to make data accessible to grassroots planners and users so that they are enabled and equipped with vital information and data that is needed for micro level planning. While it is interesting to note that as per Census 2011, there are 15.54 Million households (6.3% of total population) in India who have computers/laptops in their homes with 7.6 Million households having internet connection and as many as 14.50 million broadband subscribers as per the Telephone regulatory Authority of India (TRAI), it is still a challenge

to reach out to planners and users at the grassroots level. It is envisaged that the Data Dissemination Centres at the various Directorates across the country will play a critical role in forming the link between the available pool of data and users and in ensuring that data collected during the Census can be accessed from any part of the country. In India's experience, technology has been a boon in the process of data processing and analysis, given the volume of data that the Census operation generates, and has aided in better and quicker dissemination of data. In order to effectively disseminate census data, a combination of the latest technology incorporated into the traditional methods of dissemination of data will pave the way in ensuring easy accessibility of data to all sections of users and reach a wider audience, including the users at the grassroots level.

“Only used statistical information is useful statistical information - Adage”

Reference:

1. ***Census in India*** – A Presentation by ORGI at the United Nations Workshop for South Asian countries on Collection and Dissemination of Socio-economic Data from Population and Housing Censuses New Delhi, India 28 – 31 May 2012.
2. ***Provisional Population Totals-Paper 1 of 2011, Dr C. Chandramouli, Registrar General & Census Commissioner, India.***