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

SESSION 8: Demands for and provision of Census Micro Data

DEMANDS FOR AND PROVISION OF CENSUS MICRO DATA

Prepared by

Mr. Assefa Negera Etana
Central Statistical Agency
Ethiopia

* This document is being reproduced without formal editing.

Demands for and Provision of Census Micro Data

The Ethiopian census provides information on the number of people living in a given Region, Zone, *Wereda* (district) and *Kebele* (lowest administrative region). The census provides answer to such questions as: how many children are there in the country or region; how many are of school age; how many women are there in a given *kebele*; how many are old enough to vote; what kind of jobs they are doing; what is the level of education in a given *wereda*. The census is useful to provide these information and many others. Based on the Ethiopian experience, report on the Demand for census micro data is presented on the following issues:

1. **Users of Micro Data**

To ensure the fullest possible utilization of census results by users, a detailed census analysis plan was produced by the CTC based on the experience from previous censuses and taking into account the tabulation plan it had developed before. The target audience were not only, national, regional, zonal, and *wereda* administrations, but also, academic researchers, international agencies, and other users. Providing researcher access to micro data can help them to play important role in stimulating policy analysis and debate and assessing the effectiveness of government programmes. This requires access to good-quality statistical data if their analyses are to be effective. If they do not have access to relevant official statistical data, they will often seek to collect their own data. CSA provides sample census micro data based on the given legislation of the country.

1. Appropriate Level of Micro Data Sampling

As part of its dissemination plan, CSA has provided on line one percent of the 2007 Census micro data for public use free of charge. Each micro data file is stratified sample of the population which was created by sub sampling the full census sample that received census long form questionnaire. Initial sampling was done address-by-address in order to allow the study of family relationship and housing unit characteristics for occupied and vacant units. Sampling of people in institutions and other group quarters was done on a person-by-person basis.

2. Micro Data Confidentiality

To ensure confidentiality the micro data was annonymized, so that individuals would not be identified by researchers during the manipulation of data. CSA must maintain the trust of respondents if they are to continue to cooperate in data collection stage. Confidentiality protection is the key element of that trust. If respondents believe or perceive that CSA will not protect the confidentiality of their data, they are less likely to cooperate or provide accurate data. Therefore, information at individual level can not be divulged and such information is strictly confidential by removing direct identifiers (name, address, ID number, business name etc.) before the release of micro data file.

3. Archiving Micro Data

In the course of census operations vast amounts of data and information are generated, including sample frames, editing and tabulation programmes, statistical and analytical reports, administrative and methodological documents, and maps. These assets represent a considerable investment and comprise valuable and irreplaceable resources. Archiving encompasses actions undertaken to ensure that completed questionnaires, control forms, manuals and other relevant documents, data and accompanying metadata collected are managed and maintained in formats that that sustains their utility for current and future uses. CSA has taken measure to ensure that all these are reserved in paper and digital formats. Enough fire-proof safes had been purchased by CSA for this purpose.

Census digital data are also vulnerable to media failure, degradations and damages. Computers components and media can physically fail due to human error, natural events, infrastructure failure, malicious destruction and even just the passing of time. Taking this into account, CSA had taken measures to increase the longevity of data. Information in digital form is also at risk of being inaccessible or loss from obsolescence of technologies. Hence, as technology upon which digital content relies is replaced by a new one, effective upgrading or archiving in formats such as ASCII could help manage the impacts due to obsolescence.

Other considerations for minimizing loss of and maximizing longevity of data include issues related to data handling protocols, data storage, data security, disaster recovery and emergency

rescue of digital content. In addition to technological concerns, CSA has a backup and data security policy and procedure which are embedded in its ICT Policy.

Documentation is also an essential component of archiving that is critical for ensuring that census data and information are preserved and remain accessible in the long-term. Metadata explain the content and structure of data, provide contextual and explanatory information necessary for proper understanding and using data, and facilitate the exchange, archiving and preservation of data.

CSA has two data backup procedures. One is Disk-to-Disk backup which is taking backup from working to a storage server; and the other is from sever to other storage media devices, such as CD or DVD for application software, Plug in external tape drive(size ranges 50 to 500GB) for data, Networked (using iSCSI connection) HP 1/8 G2 Tape Autoloader for data, and Networked Dell PowerVault MD3000i for data Storage.

Recently, CSA has acquired a data backup infrastructure, as part of its ICT system infrastructure improvement. This will ensure proper documentation and avoid loss of backup data due to natural disasters and other causes.