

Data Processing's Role and Issues in the 2010 Round of Population and Housing Censuses: The case of Ghana

KB Danso-Manu, Data Processing Manager, Ghana Statistical Service, Accra,
e-mail: kb.dansomanu@statsghana.gov.gh

and

Samilia E. Mintah, Senior Statistician/Programmer, Ghana Statistical Service, Accra
e-mail: samilia.mintah@statsghana.gov.gh

Abstract

This paper is concerned with data processing issues associated with population censuses in Sub-Saharan African countries. We use Ghana's 2010 population and housing census to identify problems and challenges that arise if data processing is not planned as a holistic part of a population census. We identify appropriate technologies that enhance data processing and the various stages of censuses that have to be coordinated and incorporated with the data processing stage in the planning of national population census. These include demarcation, editing, enumeration, funding, scanning and structural editing. Our key findings are that data processing should be incorporated as an integral part into the planning of national population censuses and funds should be made available on time for the census technology to be properly tested. It is high-time Ghana considers a digital census so as to avoid problems of processing paper-based censuses. This is the more reason why data processing should be part of initial planning of censuses. Based on our findings for Ghana, we conclude the paper with recommendations for Ghana and these recommendations have heuristic value for other Sub-Saharan African countries.

Keywords: demarcation, data cleaning, enumeration, funding, scanning, structural editing.

Introduction

Population is at the centre of all planning activities. There cannot be any meaningful developmental activity which does not first consider the characteristics of the population for whom the activity is targeted at. The size of the population, its distribution over space, growth and change over time, in addition to socio-economic and housing characteristics are all important in development planning. It is the primary source of information about the number and characteristics of a given population in each small geographical unit. It takes stock of the most important asset of countries - their human capital. Population censuses are therefore critical planning tools that help policymakers plan for the future in terms of schools, clinics and hospitals, roads, urban infrastructure and more.

Accurate statistics help policymakers make the best decisions. The result from the census has to be released on timely basis; it has to be reliable, relevant and internationally comparable to form the foundation for policy and program development.

A national population census requires mapping the entire country, figuring out what technologies should be employed, mobilizing and training groups of field enumerators and data processing assistants, conducting a major public educational campaign, canvassing and listing all structures in the country, collecting household and individual information, compiling hundreds of thousands or millions of completed questionnaires, monitoring procedures for capture and processing of results, and analyzing and disseminating the data.

The data processing stage of a census is considered as the output of the census exercise and for that matter it is greatly influenced by all inputs such as cartographic work, questionnaire and manual development, recruitment and training of field workers and data collection exercise. Despite the importance of the data processing stage, most countries in the Sub-Sahara Africa do not plan adequately for its implementation and most of the time, run into avoidable challenges and delays in the release of the results.

This paper identifies and discusses the main data processing issues which affected the early release of the results of the 2010 Population and Housing Census of Ghana. It is organized into two sections. Section one examines the preparation and selection of appropriate technology for capturing the census forms. The stages in data processing such as Manual Coding, Form Preparation, Scanning, Key Verification, Data Conversion, Data Validation and Structural/Consistency Edits are examined in section two.

We conclude the paper by making recommendations for the next round of Population and Housing Censuses (PHC) in Ghana and other Sub-Saharan-African countries.

Section 1: Census Preparation

Preparatory activities for the 2010 Population and Housing Census of Ghana started with the development of a project document, budget, work plan and demarcation exercise. These were followed by the choice of technology to deploy and the development of census instruments.

New topics introduced in the 2010 PHC include:

- a) **Disability** - Any serious disability that limits a person's full participation in life activities (such as mobility, work, social life, etc.)
- b) **ICT** – ownership and access of mobile phone, internet facility for all persons 12 years and older,
 - Ownership of fixed telephone and computer by members of households
- c) **Agriculture** - Involvement of household in any agricultural activity and type of activity.

An innovation on the Ghana census questionnaire was the inclusion of household contact phone numbers which helped during the manual coding of occupation and industrial in the office. In terms of content, the 2010 census was based on the 2000 census questions apart from the new modules listed above. This demonstrates the continuity associated with censuses in Ghana.

1.1 Funding Census Operations.

Funding for the 2010 Ghana census was provided for mainly by the Government of Ghana with support from multi-lateral and bi-lateral agencies, coordinated by United Nations Population Fund (UNFPA). The initial census budget was estimated at US\$ 50 million. The Government of Ghana provided about 70% of the total cost while the remaining was provided by our Development Partners (DP).

The relationship between African governments and their development partners in funding census can best be described as a cat-and-mouse game. The DPs expect the commitment from the government to be fulfilled before making good on their commitments. Governments, however, are strapped for cash and therefore are often unable to take the lead in funding a census.

One major challenge facing censuses in sub-Sahara Africa remains the delay in releasing funds for project activities. The delay in committing funds to census activities by the development partners stem from the fact that they do not want their money to be wasted. So they tend to wait for government's commitment before making any offers. Also, some of the DPs would sponsor a specific activity on conditions that these goods or services are procured from their countries. Another reason for the lack or delay of funds could be due to donor fatigue.

Scarcity of funds underlies the late release of funds by governments for census. The late release of funds is often characterized by the perception by some departments of governments that censuses are unimportant relative to other competing national issues (such as the general election in Ghana's case). Such perceptions cause lengthy delays, increases the overall cost of the census and ultimately affects the reliability of the data produced.

Because of funding problems, field demarcation, in Ghana, could not start with the full complement of field teams. More importantly, data capture of the main census forms started seven months after field enumeration. The reason for this is that data processing seemed to be an

after-thought of planning the census and in the context of scarcity of funds; it was left on the back burner.

In the context of this funding quagmire, National Statistical Offices (NSOs) in Sub-Saharan Africa should start looking for new sources of funding for census activities by bringing on board the private sector, the commercial banks, real estate developers, manufacturing organizations, etc. Alternatively, census data should be seen as a commodity for sale to end-users in an effort to cost recovery of the overhead of generating such data. These can be practical alternatives to government subventions serving as a large source of funds for conducting national censuses. Dealing with the funding problem will ensure that data processing will be integrated into the planning of national censuses. Census requires planning and is expensive. There is no point in conducting one when the full benefit of a census cannot be achieved.

1.2 Demarcation.

Cartographic work is the largest pre-enumeration operation and its success is vital for conducting a good quality census. In order to minimize any census coverage errors, the census enumeration areas (EAs) delineated for the census fieldwork should be well-defined and should have non-overlapping boundaries. In addition, EA demarcation should all together cover the total land area of a country.

Mapping technology has made great strides over the past decades. It has moved from an activity depending on field exploration and manual drawing, to one using remote sensing and computer-assisted map management. Our intention was to acquire high resolution satellite imagery for the whole of Ghana to act as an accurate and current geographic base from which census mapping operations can be initiated and implemented. However, due to lack of funding, we were able to obtain satellite imagery for only the coastal areas of the country. As a result, the mapping exercise combined the manual and digital census mapping techniques. A total of 37,641 EAs were demarcated for the 2010 Population and Housing Census of Ghana.

Some of the challenges in demarcation were (i) the late acquisition of cartographic equipment to process the field returns, (ii) insufficient number of field teams to complete the work on schedule (iii) inadequate supervision of field work (iv) inability to acquire satellite imagery for the whole country, and (v) issues of boundary disputes in some districts.

1.3 Choice of Technology for Data Processing

When considering the technological options for data processing, we were faced with a number of questions. Some of these were:

- How to make an informed choice of the appropriate technology;
- How to maintain the integrity of the existing statistical and census systems;
- How to deal with the option of outsourcing, and management of outsourced tasks;
- How to address confidentiality concerns relating to the preferred solutions; and
- How to avoid data capture challenges encountered during the previous census.

The objectives in choosing the technology were:

- To own the system and use it for other statistical works after the census;
- To streamline the processes of receiving, grooming, data capture, validation, editing and tabulation; and
- To avoid being tied down to long term consultancy.

The Ghana Census chose the Cardiff Teleform systems using the Optical Character Recognition and Intelligent Character Recognition (OCR/ICR) scanning technology. A local Contractor was contracted to provide the service. Procurement of data processing equipment was late. The technology was therefore not tested during the Trial Census or before the main field data collection exercise. Data Processing Staff did not receive adequate training to own the data processing system. The process of scanning, verification and conversion was outsourced to the local company. There was a plan to train the Census Data Processing staff well before the census, but because of inadequate funding, the team did not have the planned training. As a result, the office is yet to reap the full benefits of the investment made in acquiring the technology.

Section 2: Census Data Processing stages & Challenges

This section describes data processing stages and the challenges we faced once funding became available for this activity. The data processing stage covered receipt of questionnaires from the field, manual editing of questionnaires, questionnaire preparation, scanning, key verification, conversion, data reformatting, editing and tabulation.

Two main questionnaires (PHC-1A and PHC-1B) were used for the enumeration of household and non-household (group quarters) populations respectively during the 2010 Ghana Population and Housing Census. In addition to these, two summary sheets were used namely: Enumeration Results Sheet (PHC3) and Final Summary Sheets (PHC4). The PHC3 contained summary information on an Enumeration Area, the smallest geographical unit assigned to each Enumerator whilst the PHC4 was the summary information on localities and the availability of facilities like telecommunication, health, educational and toilet in the localities.

Information from the PHC3 was collated to release the provisional results whilst PHC1A and PHC1B were processed to release the main census results.

Fijitsu fi-5900C and fi-6130 scanners and Teleform version 10.4 software with Optical Character Recognition and Intelligent Character Recognition (OCR/ICR) scanning technology were used in processing the 2010 census. The process of scanning, verification and conversion was outsourced to a local company. The Ghana 2010 Population and Housing Census was generally successful but some of the challenges encountered are identified and discussed below.

2.1 Storage and Office Space

At the end of the field enumeration, questionnaires and forms were packed together with field materials into boxes at the district offices. They were then dispatched to the regional offices and finally to the national census secretariat.

One of the main challenges encountered was the issue of inadequate storage space for the census questionnaires. The capacity of the available storage facility could contain only about 20 percent of the total number of boxes received. As a result all the offices of the census secretariat were used as a temporary storage facility. However, the total office spaces made available was not sufficient to store all the boxes. The content of the boxes were therefore transferred into satchels so as to create more space. In the process questionnaires from different Enumeration Areas got mixed up. There was a plan to build a storage warehouse for the census operations, but this never materialised.

Accommodation for editing and coding staff was also affected and this caused disruption in the flow of work. The GSS research library was used to accommodate the manual editing and coding staff. Later, the place was used as additional scanning rooms.

2.2 Recruitment and Training of Data Processing Assistants

An application form was designed and uploaded onto the Statistical Service website. Interested applicants had to fill out and submit completed forms for consideration. Applicants, who met the requirements, were shortlisted and interviewed. These persons were successfully trained in batches and assigned. At a point in time there were over one thousand two hundred temporary data processing assistants working 3 x 8 hourly shifts per day and 7 days a week. Management of the large number of temporary staff was another challenge. However, strategies were developed to control the situation and all efforts were made to discourage absenteeism.

2.3 Badly Printed and Badly Handled Questionnaires

The technology used had its standard paper requirements. Unfortunately, some of the questionnaires used for enumeration did not meet the specified requirements. The delay in procuring data processing technology also affected the printing of questionnaires. Due to the short period given to the printing contractor, questionnaires delivered were not enough to complete field enumeration during the specified period of two weeks. As a result, there were shortages of questionnaires in most part of the country. Additional printers were contracted to print some of the questionnaires. In their rush, most of the printed questionnaires did not meet the specification given.

Some of the problems encountered during the data processing preparatory activities were as follows:

- Bigger or shorter than the normal A4 size required ('long sleeve' or 'short sleeve') forms;
- Chopped off reference blocks (triangles at the corners of the census forms technically used by the scanner to align the forms);
- Faded reference blocks;
- Pages turned up-side down, missing pages and bad paper quality;
- Bad handling of questionnaires by field enumerators.

New questionnaires printed in-house were used for the transcription, thus creating additional cost in paper, time and consumables. About a third of the 6 million questionnaires were estimated to be defective.

2.4 Delays in the Procurement of Data Processing Technology

Procurement of data processing equipment was late. The initial bidding and evaluation processes were cancelled because our Development Partners guidelines were not followed. The procurement process was re-started and it took a long time to complete. The technology was neither tested during the Trial Census nor before the main field data collection exercise. Data Processing Staff did not receive adequate training to own the data processing system. The process of scanning, verification and conversion was outsourced to a local contractor.

2.5 Manual Editing and Coding

The 2010 Population and Housing census had no plan for office editing and coding. It had been ruled out of the census process. All editing and coding were to be done in the field and the field supervisors were to effectively and efficiently execute this task. In this regard budgetary allocations were not made for office editing. However, there was the need to recode occupation and industry using the descriptions on the questionnaires because most of them had been misclassified. This activity delayed data processing period.

2.6 Scanning Process

At the beginning of the scanning process 62 desktop computers and six enterprise scanners (fi-5900C) were networked. Six of the computers were attached to the enterprise scanners (Scan Station), 20 were used for data reading and interpretation (Reader Module) whilst the remaining 36 were used for key verification (Verifier Module). However network failure reduced the capacity of the scanners to one questionnaire per batch instead of five. This slowed down the whole process drastically. Some questionnaire images could not be archived using the Meridio software system in place. To remedy this situation, additional equipment and personnel were deployed. Forty-six new stand-alone scanners (fi-6130) were acquired which operated outside the network. Even though this rationalisation enhanced the timely scanning of all the census questionnaires, taking backups was a tedious exercise. Images of questionnaires scanned by the standalone scanners could not be obtained. Constant power interruptions also affected the work flow and data lost had to be rescanned. The scanning process alone took about ten months (from July 2011 to April 2012) instead of the estimated six months.

2.7 Data Conversion

Data conversion was the final stage of producing ASCII data for tabulation in Census and Survey Processing System (CSPro). The main census questionnaire had 14 pages and each page was scanned into a different data file. The conversion process performed structural edits and linked all the pages of each questionnaire, household or institution together. This required certain key variables which were absent in some cases even after the key verification process.

Another process, pre-conversion was introduced to ensure that the key variables to be used for the linking were filled. This activity was done per district and so EA data from the stand alone scanners had to be merged for the pre-conversion exercise.

2.8 Structural Editing, Consistency Edits and Data Cleaning

This activity was done in CSPro and required fixed ASCII data. However, the data when converted directly from Microsoft Access to fixed ASCII had problems. To resolve this issue the Data reformatting process was introduced. By this process the data had to be first converted to SPSS for further processing such as insertion of the record types, fixing of variable lengths and then finally

reformatted to fixed ASCII in CSPro for final editing and imputations. Most of the initial tabulations were done in CSPro.

Conclusion/Recommendations

Census is the core of National Statistical Systems. It would continue to be an indispensable tool for ensuring political representation, equitable allocation of government resources and the distribution of services. It is also essential for monitoring and achieving developmental goals such as the Millennium Development Goals (MDGs) and Poverty Reduction. In order to avert funding crisis, delays in releasing final results and to ensure their future stability population and housing censuses should be incorporated into the national statistical system and not to be seen as a one-time activity by the National Statistical Offices.

Going forward, what lessons are there for Ghana and other Sub-Saharan African countries? Considering the numerous challenges discussed in this paper, do we do away with paper-based censuses and go digital? We make 10 specific recommendations here:

- 1) NSOs should start looking for new sources of funding for census activities by bringing on board the private sector, the commercial banks, real estate developers, manufacturing organizations, etc. who use population data for their activities;
- 2) Governments should not fund censuses through subventions. If they do, this should be done in a cost recovery manner where census results should be sold on a commercial basis for end users. Countries like the United States of America use such a model in funding their censuses;
- 3) Census activities should be planned well ahead of time, at least five years before the intended date. Data Processing technology should be procured earlier in order to allow proper testing of the systems,
- 4) There should be adequate documentation of previous censuses to avert same mistakes repeated from one census to another;
- 5) Research should be conducted into census costs and operational methods to determine what practical measures can be taken to reduce costs, as well as how to maximise the timely dissemination and use of census results;
- 6) There is the need to put in place an effective and efficient monitoring system during the demarcation exercise to ensure the correct sizes of EAs are maintained and every community in the country is visited during demarcation;
- 7) To resolve many of the problems encountered in data processing Ghana, and for that matter Sub-Saharan African countries, should consider reverting to a digital census so many of the issues of data processing identified in section 2 can be avoided (storage, printing, scanning, data conversion etc.). This may sound out of the box but it may be more efficient and effective than the current paper surveys that have to be translated into digital form after the census enumeration;
- 8) It is important to perform comprehensive tests of the census processes well ahead of the census date;
- 9) Fund for Census Data Processing activities should be released on time. The data processing technology to be employed should be fully tested during the trial census and data processing officers should have adequate training to efficiently manage the systems;
- 10) A system should be put in place such that training of field enumerators would be uniform throughout the country.

References

- Arij Dekker**, (2001) Adapting new technologies to census operations
Carolyn Lamere, (September 2012) Counting the World: UNFPA Highlights the Challenges of Census Taking
Ghana Statistical Service, (2012) Administrative Report on the 2010 Population and Housing Census
Ghana Statistical Service, (2010) Enumerator's Manual - Ghana 2010 Population and Housing Census
Richard Leete, (2001) Funding Crises in the 2000 Round of Population Censuses.