Challenges in Census data processing using contemporary technologies: The case of Malawi

Msosa Angela;
National Statistical Office
Technical Services
Zomba

E-mail:amsosa@statistics.gov.mw,angela_msosa@yahoo.com

Malawi is one of the countries that has recently conducted its census within this round of 2010 population and housing census. From the preliminary results, the country's population is at just over 13 million. Geographically Malawi is situated in Central Africa

The importance of population data can not be over emphasized when it comes to the country's planning needs. Census provides data by locality ,age and sex among other details. National Statistical offices(NSO), as census implementing agencies, are faced with the challenge of releasing census results timely after census enumeration. This among other things entails shortening the data capture period mostly through more improved systems such as scanning. Malawi Statistical Office has been using scanning as a data capture method for small scale surveys since 2002. However , scanning at a large scale was used for the first time with the 2008 Population and housing census.

CENSUS PREPARATIONS

Pilot for the recently done census was conducted in 2007. By the time the pilot census was done, the Malawi Statistical office had decided for the capture method to be manual entry. Before the pilot census, Malawi NSO had explored the possibility of using scanning to capture data for the planned census. However after an independent review of the capacity and resources required to use such technology for such a big census exercise, it was suggested that manual data capture would be the most appropriate. Specifically it was noted that to successfully carry out the scanning process as self sufficient institution:

- Malawi NSO would have to train sufficient ICT staff to ensure that the office has capable staff to handle the scanning operation
- Malawi NSO would have to train ICT staff in good time to subjectively self-assess the scanning requirements
- Malawi NSO would have to have build capacity to support the scanning equipment i.e scanners and related accessories
- Malawi NSO would have sufficient time to test the scanning technology to ensure that the hired enumerators will be able to follow the requirements of data collection using the scannable forms
- Malawi NSO would have to secure sufficient financial resources to procure the scanning equipment and software

Even though Malawi NSO had used scanning technology before, it was at quite a low scale such that it was impractical to assume that the same success would be achieved for a census exercise. Teleform and Eyes and Hands are the 2 softwares that had been used by Malawi NSO before. They happen to be OMR technology as well. Only one out of all the surveys that NSO conducts uses scanning as a capture method. Teleform version 7.1 was used as the first scanning system and it was the only time it was used. It was found to be a not so easy software to support in terms of initial system set up and form design. NSO had to rely heavily on external technical assistance to run the system. As a result the subsequent survey was captured using Eyes and Hands(E&H) software. This software was more manageable in terms of form design and system set up including printing of forms .However, looking at countries that had used scanning in the region, none had used the E&H software. As such, the pilot census was captured manually in readiness for the main census exercise.

After the pilot census in 2007, Malawi NSO was ready to carry out the main census exercise in September 2008. However, NSO had to shift the schedule to conduct the exercise in June 2008 as there was an urgent need for population figures. To achieve this, NSO decided to revise the data processing plan and use scanning . This meant that a system had to be put in place promptly to fast track the census process including data capture. DRS systems LTD which is a UK based company was identified which specializes in Optimal Mark Reader(OMR) scanning technology. As such the following was the approach:

- A single sourcing to identify the scanning company was used. DRS company was to assist NSO with census scanning based on the company's previous experiences in some African countries. DRS was contracted to assist with the whole data capture process including supply of the scanning equipment, census forms and assist in training of enumerators and data processing staff.

- A quick pretest was done for a few households to test the scannable census form . The pretest was also aimed at assessing the training requirements of the enemerators. Following the pretest enumeration, the forms were scanned

SCANNING OPERATION

For the main census, 6 scanners that were supplied by DRS Ltd were used for scanning. The staffing was as follows:

- 120 form checking staff with 4 supervisors
- 12 scanner operators with 2 supervisors
- 90 key correction staff , that were verifying the scanned data on 45 PCs with 4 supervisors

Scanning was done in about 10 weeks for 3.5 million forms that had to be split into two resulting in nearly 7 million forms being scanned. Key corrections for almost 80 percent of the forms took about 18 weeks.

Problems encountered during scanning included:

- Slowness of the DRS system especially the key correction program. It was difficult to know whether it was the DRS system or the computer network that was slowing the system
- Breakdown of some scanners which requires hardware support that was not available on site though it was agreed in the contract with the scanner suppliers that hardware support would be available at all times of the scanning process
- Forms being rejected by the scanner, in some cases due to poor form handling resulting in rejection. In one instance an entire batch of forms was rejected by the scanner even though the form looked clean enough implying that it could have been a problem with the quality of the form. It was difficult though to establish/ confirm

After the capture was complete, Malawi NSO proceeded with the data processing stage of editing. It was at this stage the scanned data showed a few problems which could have easily been avoided at the scanning stage. These included:

- Missing data on the form as the scanner couldn't read the marks on the forms. This was both due to poor handling by the enumerators and office staff. Even though office staff had checked each of the marks on the forms, it was found that that process didn't improve the missing marks It was clear that the human eye was not able to detect errors on forms designed to be read by a machine
- Extra data was found in the household form due to false marks read by the scanner. Sometimes extra data was introduced at the key correction stage though it was minimal
- Since before scanning a batch, a batch header was scanned to attach geocodes to the scanned data of the batch, it was found that some batch geocodes did not match any of the geocodes available in the master census geocode list. This was because the batch geocodes were manually entered as DRS had not prepared scannable batch header forms to errors introduced by manual entry.
- The scanning system was not so strict in disallowing form geocodes that didn't match the batch headers geocodes. A batch header geocode was not distinct to one EA. This could have also arisen during the office checking stage where forms from different batches got mixed up

LESSONS LEARNT

It was clear at the editing stage that the process of setting up the master geocode list in a data capture system helps in introducing foreign geocodes in the data. Since during a scanning operation, once forms get in the office, they are prepared to sit on the machine to be read, it is assumed that the marks including the geocodes on the form will be read .

This assumption is quite fatal since if the proportion of poorly marked forms is large, the time it takes to sort out the expected geocodes once scanning is complete could get longer.

It was also clear that right from the field work stage, some supervisors were not thoroughly checking the forms of the enumerator. Otherwise if an enumerator was found to be improperly filling the form or poorly handling the form, then they ought to have been advised to improve. This could have been due to limited knowledge by the supervisors due to limited training and experience on scannable forms. Consequently, a high number of staff was hired to manually check forms, though not effective. This is contrary to expectations that with scanning, very few data processing staff is required .

It was also observed that if detailed reports from the DRS were available, they could have assisted with the 'key correction ' exercise. It was found that during key corrections marks on age and sex which are key census variables,could have been corrected. However in the absence of appropriate monitoring reports , it was difficult to pick up the fact that keyers were not paying attention to these 2 fields.

In some instances, since the support was being remotely offered from UK where the company is based, it was difficult to get prompt assistance if a problem was found with the system. It was also noted that once the major component of the contract was done, i.e scanning, getting prompt response on other queries to do with the structure of scanned data was not as prompt. It was noted that at the editing stage, it is equally important for the scanning company to be around for queries to do with transfer of data from their system as it was noted in Malawi's case.

In summary, the data capture period was drastically reduced with scanning. However, it took a lot of resources, both human and financial, to achieve the required results