Modernising data collection for Household Surveys
Luxolo Lengs  (Statistics South Africa)
luxolol@statssa.gov.za

Abstract
Traditionally we’ve conducted our surveys using paper as our main data collection instrument. 2 years ago we took a strategic decision to introduce digital collection methods. This decision taken after a few years of testing various hardware and software technologies and evaluating what infrastructure would suit such a change.
In the last 12 months we’ve conducted 2 surveys Citizen Satisfaction Survey (22 000 households) and a Community Survey (1.5 million households) using 300 and 14 000 Android tablets respectively. Our vision for both surveys was to complete the surveys with digital solutions and not have any paper in the value chain.

Citizen Satisfaction Survey
This survey was conducted in late 2015 and was used as a pilot to test and verify the readiness of the different facets of the statistical value chain for the planned Community Survey which would be executed in early 2016. The following items were digitised and paperless:
• Human Resources processes (Recruitment and Sourcing & Selection)
• Dwelling frame which was used as our database from which households were sampled
• Training and competency assessments
• Project monitoring
The project was concluded within 3 months and we met our expectations to digitise the processes.

Community Survey 2016
This survey was conducted in early 2016 and had a total coverage of 1.5 million households. Similar to the Citizen Satisfaction Survey, we digitised all processes.
The major change with this survey, apart from digitisation, was the use of an Access Point Network (APN) which gave us a private segment of the Telco’s mobile network. All data transmission from the enumerator’s tablet was encrypted and only decrypted at our data centre. The data was channelled through the encrypted network to a private cloud which was hosted offsite but connected to our data
centre. This gave us end to end control and visibility across the various handshakes (tablet to telco’s network to data centre) that the data would flow through.

Another major improvement added to this project was a private navigation app that guided our 12 000 enumerators to the many households they needed to collect information from. The enumerators would only see the dwelling units assigned to them on their navigation app. Because each household in the project had a set of unique coordinates, we could then use the data from the Navigation solution to verify if the Enumerator had gone to the household – this allowed us to monitor down to sub-meter level if the correct households we enumerated.

We also deployed a Mobile Device Management (MDM) solution for the 14 000 Android devices to remotely manage any issues with the devices. If there were any changes to contents (training material updates, maps or software patches) then the MDM solution would be used to send commands to the devices to update/refresh. Benefits: We used 60% of the allocated budget for the project and managed to complete it within 3 months. Ten years earlier, it took us a year to complete the same survey and we used double the budget used this year while only visiting 250 000 households then.
Modernising Data Collection for Household Surveys by Statistics South Africa

(Author: Denzyl Papier & Luxolo Lengs)
# Contents

1. Summary .......................................................................................................................... 3
2. Introduction ....................................................................................................................... 4
3. Objective .......................................................................................................................... 5
4. Approach and Outcomes ................................................................................................. 6
   4.1 Citizen Satisfactory Survey (CSS) ........................................................................... 6
   4.2 Community Survey 2016 .......................................................................................... 8
5. Looking forward (Way Forward) .................................................................................... 9
6. Conclusion ....................................................................................................................... 10
1. Summary
Traditionally we’ve conducted our surveys using paper as our main data collection instrument. 2 years ago we took a strategic decision to introduce digital collection methods. This decision taken after a few years of testing various hardware and software technologies and evaluating what infrastructure would suit such a change.

During 2015 -2016 Statistics South Africa (Stats SA) conducted 2 surveys:

1. Citizen Satisfaction Survey (22 000 households); and
2. Community Survey (1.5 million households) using 300 and 14 000 Android tablets respectively.

Our vision for both surveys was to complete the surveys with digital solutions and not have any paper in the value chain.
2. Introduction

Statistics South Africa (Stats SA) is mandated by the Statistics Act No.6 of 1999 to conduct quinquennial censuses for the country. Since the country’s independence in 1994, Stats SA has undertaken three population censuses in 1996, 2001 and 2011. These censuses have generated diverse demographic and socio-economic information that has guided the formulation of policies and interventions aimed at enhancing the development of South Africa. The results have also provided valuable insights into the progress South Africa has made in achieving the priority areas of the Medium Term Strategic Framework (MTSF 2009-2014), the Millennium Development Goals (MDGs) and the National Development Plan (NDP).

Stats SA was supposed to undertake the next census in 2016 but was cancelled due to lack of enough funding. A decision was then taken to use the available funds and conduct a large scale survey, Community Survey 2016 (CS 2016). The main objective of CS 2016 is to provide an estimate of the country’s population at local municipality level and data that will serve as input into the municipalities’ Integrated Development Plans (IDP) and infrastructure investment budgeting.

Unlike the previous censuses and surveys, CS 2016 will be a technologically driven survey. All major processes in the CS 2016 value chain will be technologically driven including advocacy, publicity, recruitment, training, enumeration, supervision, monitoring, progress reporting, data processing and dissemination.

One of the key activities of the CS 2016 is enumeration. It is envisaged that enumeration will be conducted for a period of six weeks from March to mid-April 2016 using digital devices as data collection tools. Digital devices will also be used for remote supervision, monitoring and real time progress reporting. Implementing technology in all CS 2016 processes will significantly reduce the cost of field operations and improve the quality of data that will be collected and the results will be released timely.
3. Objective

The primary objective was to use technology and geospatial information to provide a much faster, more efficient and better quality product: an estimate of the population count and the number of households in a Municipality in order to provide data that will serve as input into their Integrated Development plans (IDP) and infrastructure investment budgeting. The following was measured:

- Estimate of Household count by Municipalities
- Estimate of Population count by Municipality
- Proportional changes of services delivery variables

The total number of enumeration areas (EAs) was 103 576 and the frame used the Census 2011 frame. More details on the sample are listed on the table below. CS 2016 tests was conducted in October 2015 followed by field operations in March 2016.

Below is a high level view of key milestones.

*Figure 1: Key Milestones*
4. Approach and Outcomes

4.1 Citizen Satisfactory Survey (CSS)

This survey was conducted in late 2015 and was used as a pilot to test and verify the readiness of the different facets of the statistical value chain for the planned Community Survey which would be executed in early 2016. The following items were digitised and paperless:

- **Human Resources processes (Recruitment and Sourcing & Selection)**

  The recruitment process was done in three levels – a. normal recruitment for Monitoring & Evaluation monitors, b. Database recruitment for fieldworkers and c. targeted recruitment (done parallel to the normal recruitment for areas that have already been identified as problematic).

  The methodology for the recruitment of field staff has moved from the traditional manual process to an online one where applications were done on a tablet, smartphone or computer which had internet access. Recruitment for field staff was decentralized to provincial and district offices supported by a centralized control structure at Head Office.

  All recruitment activities were carried out by Human Resources Officers (HRO) at Head Office, Provincial and District Offices and progress was monitored on regular basis through the Human Resources Management Team. An online recruitment system was developed by Business Modernisation (BM) with specifications as outlined by HRM.

  Below is a high level view of the Infrastructure used to support the Human Resource (HR) recruitment system.

![Figure 2: Infrastructure for HR and Training](image-url)
• **Training and competency assessments**

The use of on-line recruitment minimised nepotism and other human related issues and errors i.e. no human intervention and allows accountability. The system was able to administer large number of applicants. The selection process was not biased as it didn’t rely on human preferences *(refer to 4.1).*

The system (LMS) was successfully utilised during training as well as appointment of field staff as it provided the scores of the trainees immediately to human resource management for selection.

• **Dwelling frame**

The Dwelling Frame is made up of all structures in the country, including dwelling units and it was used for sampling of dwelling units within each field worker unit countrywide.

A web based tool was used, nationally, to update the Dwelling Frame, draw the sample and create field worker units and FWSUs for fieldwork. The updating process was office-based and it used a web-based application tool to spatially capture Census 2011 listings.

![Dwelling Frame before update](image1.png) ![Dwelling Frame post update](image2.png)

*Figure 3: Dwelling Frame Update*

• **Project monitoring**

Using a business intelligence application, SASBI, dashboards was used for daily progress report. It provided daily reporting with colour coded spatial display of progress per province, municipality and district.
Supervisors could manage, intervene and track progress for the field worker units, under their supervision. The technology allowed for access to anyone, given rights and involved in the project, through smart phones and tablets.

The project (CSS) was concluded within 3 months and we met our expectations to digitise the processes.

4.2 Community Survey 2016

This survey was conducted in early 2016 and had a total coverage of 1.5 million households. Similar to the Citizen Satisfaction Survey, we digitised all processes.

The major change with this survey, apart from digitisation, was the use of an Access Point Network (APN) which gave us a private segment of the Telco’s mobile network. All data transmission from the enumerator’s tablet was encrypted and only decrypted at our data centre. The data was channelled through the encrypted network to a private cloud which was hosted offsite but connected to our data centre. This gave us end to end control and visibility across the various handshakes (tablet to telco’s network to data centre) that the data would flow through – as depicted below in Figure 4.

Another major improvement added to this project was a private navigation app that guided our 12 000 enumerators to the many households they needed to collect information from. The enumerators would only see the dwelling units assigned to them on their navigation app. Because each household in the project had a set of unique coordinates, we could then use the data from the Navigation solution to verify if the Enumerator had gone to the household – this allowed us to monitor down to sub-meter level if the correct households we enumerated.

We also deployed a Mobile Device Management (MDM) solution for the 14 000 Android devices to remotely manage any issues with the devices. If there were any changes to contents (training material updates, maps or software patches) then the MDM solution would be used to send commands to the devices to update/refresh.
Successes:

- Collected data digitally for 2,7million questionnaires in 47 days.
- Accurate assignment control ensured proper survey management.
- Each questionnaire was GPS tagged during DC and ring-fenced to sampled point
- Data was handed over to the content and products team when the last day of mop up was complete.

Benefits: We used 60% of the allocated budget for the project and managed to complete it within 3 months. Ten years earlier, it took us a year to complete the same survey and we used double the budget used this year while only visiting 250 000 households then.

5. Looking forward (Way Forward)

The two surveys mentioned were game changing surveys for us. They ushered in a new way for data collection and allowed us to start taking a look to what might be other and Smarter Data Sourcing opportunities.

Going forward, we intend on digitising the rest of our surveys so that the same benefits can be realised across all of our surveys and Censuses.

We’ve also started to look at other ways of improving our data collection processes. In the last 2 years we’ve also started looking at how we can use scanner data from check-outs at supermarkets as input into our Economic surveys such as Price Inflation statistics. Most of the products that are within the basket of goods measured as input for price statistics are sold at supermarkets so this makes scanner data an attractive proposition.
We are also researching the use of social media analytics with Big Data technologies to test if there might be benefits that can be derived from such data. While Big data is an appealing proposition, there are still some issues (technical and governance) that need to be addressed before it can be used as a source for official statistics.

6. Conclusion
In conclusion, the following:

- Device-based data collection reduced processing time and improved quality
- Significant cost savings in relation to printing of questionnaires, forward/reverse logistics etc.
- Technology is only an enabler to data collection, core monitoring procedures and primary methodological processes need to be refined accordingly.
- Don’t rush, the environment must be right