

Web surveys for offline rural communities

Beatrice W. Gichohi

*Chief Manager Research Planning and Development, Tana and Athi Rivers Development Authority (TARDA),
Nairobi, Kenya*

E-mail: bettygichohi@gmail.com

Abstract. Use of internet based tools of data collection is a global trend that is rapidly gaining popularity over the traditional paper and pencil method. A typical web survey will be self-administered by a literate community, with access to internet. Even though technology is one of the fastest growing sectors in Kenya, with a particularly rapid increase in access to the internet and mobile phones, web surveys are yet to gain popularity as the data collection method of choice. This is mostly attributable to low literacy levels and poor access to internet connectivity in the rural areas, where the biggest proportion of population lives. Motivated by the inherent benefits associated with use of internet based surveys, one organization in Kenya, TARDA, successfully developed and adapted web based tools to collect data from 3,400 households most of who were offline. The mobile phone was the basic tool of data collection. The survey yielded a response rate of 96.5%, was completed within one week, reduced cost of data collection by 60% and eliminated errors resulting from data entry. The paper looks at the design and execution of the survey, the success story and lessons learned from the challenges encountered.

Keywords: Internet, respondents, household, mobile phone

1. Introduction

TARDA collaborated with stakeholders to formulate a 20-year Integrated Regional Development Plan (IRDP) for the Tana and Athi River Basins. Developing the IRDP entailed conducting studies to understand development priorities from the perspective of the various stakeholders. Data on multi-sectored thematic areas including economic, socio-cultural, natural resources, infrastructure, population and settlement, and institutions was collected. A Web based tool adapted to suit the characteristics of the study population was used to administer the survey, with the mobile phone as the basic tool of data collection.

Use of technology in data collection is rapidly gaining popularity. Web surveys have the ability to collect large quantities of data, at a low cost and over a short period of time. Online surveys have been known to work best when the target population is available online and is literate enough to self-administer a questionnaire. According to the Communications Commission of Kenya (CCK) 2011–2012 Report, internet usage grew by 19.2% to 7 million users, while mobile phone usage grew by 1.7% to 29.7 million users, in Kenya. However, despite the positive outcomes that

high speed internet has brought to Kenya's economy, disparity in access remains a large obstacle as majority of internet users are located in the capital city, Nairobi. About 72% of Kenyans lack internet access, while 11% lack telephone access. The most affected are communities living in the rural areas, who do not have access to the internet.

The 2009 KNBS Census reported a population of 14.3 Million people living in the Tana and Athi River Basins, which was about 37% of Kenya's population. 77% of this population lived in the rural areas which are characterized by low access to internet and low levels of literacy. TARDA developed and adapted a web survey to collect data from respondents mostly drawn from rural Kenya. The data collection tool was a hybrid of the online web surveys and the conventional enumerator-assisted data collection methods. This paper looks at how the web survey administered by field enumerators was designed and used to collect data from 3,400 households.

2. Rationale

The internet has been turning around many disciplines including the survey research industry. Kenya's

communication technology is the fastest growing sector, with increased access to mobile phone and internet technology. Most mobile phone users have access to 3G internet technology. About 89% of Kenya's population has access to mobile phones. With decreasing cost of mobile phones, majority of the people today own mobile phones which are internet enabled. The mobile phone seems to be closing the technology and literacy gaps. In Kenya people with very basic education are able to use the mobile phone to execute complex tasks such as mobile money banking and transfer. TARDA therefore attempted to leverage on this growing use of mobile phones for the IRDP house hold survey.

Previous experience has shown that if used well online methods of data collection are more cost effective, yield cleaner data and get faster results. Working on a limited budget, the organization had to devise a method of collecting data from a sample of 3,400 household drawn from all the 19 counties in the region.

TARDA was also trying to assess the ability and readiness of the stakeholders to use internet and mobile gadgets to collect and share information. The organization was in the process of developing a regional Data Center, a one stop shop for data on development indicators. Running and maintaining a data center requires regular updating of data. In using the mobile phone to collect data the enumerators were able to demonstrate the ability and readiness of the people to use mobile gadgets to perform other tasks besides communication and money transfer.

3. Methodology

The Tana and Athi Basins covers an area occupied by 17 counties. The area is broadly divided into three regions; upper, middle and lower catchments. The house hold survey was guided by an overall coordinator assisted by catchment coordinators.

An online questionnaire was developed. The questionnaire had been designed to capture development priorities as perceived by the community, guided by the following thematic areas of development; economic, population, socio- cultural, environmental resources, infrastructure, human settlement and institutions. The questionnaire had 120 variables, divided into seven sections of the thematic areas, and had both open ended and closed questions. A web link was created and uploaded on the organization's website.

An Average of 200 questionnaires was administered in each County. A total of 3,400 households were se-

lected for the survey. A directory obtained from the Kenya Population and Housing census 2009 was used as the sampling frame. Random sampling was used to select the households to be studied. From each county, four wards were selected using simple random sampling and from each ward a cluster sample of 50 households was selected. Each cluster was a sub village.

Catchment coordinators identified enumerators who would guide respondents through the web based questionnaire. To qualify as an enumerator a person had to be a resident of the ward, have a 3G enabled mobile phone and have at least secondary school level of education. Using members of the community to collect data members enhanced the response rate of the survey since they are able win the trust of the respondents into participating in the survey. The enumerators were trained on how to guide the respondents through the web survey and other ethical requisites in the exercise.

Bulk internet bundles were purchased and distributed to the enumerators. Each of the enumerators would visit the respondents at a pre arranged convenient location (home, their farms, shopping center and market). The enumerator would introduce himself, explain the data collection exercise and introduce the respondent to the web survey. Respondents with previous exposure to web applications would be given the opportunity to respond to the questionnaire with the guidance of the enumerator while those who didn't possess the ability to read and respond to the questions were assisted. An offline survey had been developed to be used in instances there was no access to internet. The offline copy had been saved in the phones' memory and would be used to record the responses which would be uploaded later.

Back in the office, the catchment coordinators were tracking the responses as they were being submitted from the field. Data from completed questionnaires would be downloaded at regular intervals and screened to identify any errors resulting from data collection. The enumerators would be informed of any errors in the data and they would seek clarification from the respondents.

Each enumerator would complete about 15 questionnaires each day using about 100 MBs of Internet. The data collection exercise took five days within which all data had been collected, downloaded and cleaned.

4. Achievements

High response rate: The biggest success of the web survey was the high response rate. Out of the Targeted

3,400 households, 3,284 responded translating to a response rate of 96.5%. This could be attributed to the relevance of the survey to the community needs, excitement about the web survey method and flexibility of the enumerators to meet the respondents at their convenience. Enumerators said it was easier to record open ended questions on phone.

Cost effective: The Web survey was cost effective since the only cost incurred was purchase of data bundles and compensation of the enumerators for the five days they collected data. This method saved on cost of printing the questionnaires, field travels since enumerators were drawn from the local community and data entry. The total cost of administering the web survey was about KES 300,000 (\$3500) against an initial budget of about KES 1,000,000 (\$11,600) of collecting and entering data using the conventional methods.

Clean data: This method eliminated errors resulting from data entry. The coordinators were able to download and clean the data in their offices as data collection was going on in the field. They would send their queries to the enumerators who would seek clarification from respondents and then send feedback to the office. By the end of the five days, there was clean data downloaded and ready for analysis.

Quick results: It took four days to collect, enter and clean data from 3,200 respondents. Had the conventional method of data collection using pen and paper followed by data entry been used, the exercise would have taken about 1 month.

Respondent guidance and interpretation of the questions: Respondents were guided in taking the household survey. This ensured that everybody regardless of their levels of education had the same interpretation of the questions. The questionnaire had some complex questions and it was therefore easy for the respondents to seek clarification from the enumerators. This also minimized incidents of missing responses.

5. Challenges encountered

Internet access: The initial assumption that internet would be accessible everywhere, any time was proved wrong. In a few instances Enumerators were unable to access internet. They however were able to use the offline version of the survey which they would later upload. This slowed down the pace of the work since some of the time allocated to data collection was used uploading data. There were also incidents of slow internet connection which could sometimes slow down the exercise

Data security: With rapid growth of incidents of cyber crime in Kenya, Some respondents were uncomfortable with their responses being uploaded on line. This was despite the fact that the questionnaire contained very little personal information. This was unpredictable and the data collection coordinators had to develop a solution. A few questionnaires were printed and used to collect data. To respect the respondent's option not to have their responses uploaded on internet, the questionnaires were sent to the office and responses added to the already downloaded raw data

Reliability of the mobile phone as a data collection: Some mobile phones batteries could not last the whole day. This posed a problem in some rural areas where there was no access to power. Enumerators had to devise methods of ensuring that they were able to collect data throughout the day. Solutions included carrying an extra phone and swapping SIM cards when the first phone's power gets low. There were also incidents where respondents' mobile phones were used. In future power backups should be provided.

Literacy levels and previous exposure: Literacy levels vary greatly with from one county to the other. Most of the respondents in Nairobi and some counties in the upper and middle catchments seemed to have some exposure to use of internet and were able to respond to the survey without a lot of assistance. In other counties majority of the respondents required a lot of assistance. Questionnaires should be designed to fit the varied literacy levels.

6. Conclusion

With increasing access to mobile phone technology and internet all over the world, it is increasingly becoming easier to administer web surveys to communities that would typically be classified as offline. Kenya's population is ready and enthusiastic about adoption of technology and respondents were positive about responding to a questionnaire administered through the mobile phone. They termed it as a more friendly method of data collection. The initial skepticism about using the web survey revolved around issues of low literacy levels and internet access. The response to this was to design a method which was a hybrid of the web survey and the conventional enumerator assisted data collection. This ensured that the exercise benefited from the strengths of each of these two methods. Introduction of enumerators with internet enabled phones to the exercise ensured that respondents'

literacy levels and internet access were no longer an issue in administering the survey. The questionnaires were only administered to the sampled respondents, thus minimizing the sampling error. As Kenya prepares to conduct the next population census in 2019, adapted web surveys could be considered as a tool of data collection for enhanced cost effectiveness, speed in administering the survey, accuracy of results and high response rate. However in adopting such surveys researchers should carefully study the target population and adapt to the situation. Issues of sampling, reliability of internet access and ability of the respondents to self administer a questionnaire should be considered.

References

- [1] T.M. Archer, Web-based surveys, *Journal of Extension* **41** (2003), online.
- [2] A.Z. Aster, Consumer Research Goes Online, *Marketing Magazine* (2004), 13–14.
- [3] KNBS Population and Housing census 2009, Volume 1A, Population Distribution by Administrative Units, (2010).
- [4] KNBS Population and Housing Census 2009, Volume 1B, Population Distribution by Political Units, (2010).
- [5] KNBS Population and Housing Census 2009, Volume 2, Population Distribution by Socio-Economic Characteristics, (2010)
- [6] D. Rivers, Probability-Based Web Surveying: An Overview, 2002, Paper presented at the annual conference of the American Association for Public Opinion Research, Portland, OR, May.