Ethiopian Statistical Association (ESA)

Pilot Project on the Use of Mobile Technology for Data Collection

Current status of the use of mobile technologies for data collection in Ethiopia

PROJECT TITLE: Strengthening the capacity of African countries to use mobile technologies to collect data for effective policy and decision making

By joint collaboration among the Central Statistical Agency, the Ethiopian Statistical Association and African Center for Statistics, UNECA

November 2014

Table of Contents

Acronyms

CSA	Central Statistical Agency of Ethiopia
AgPPI	Agricultural Producer Price Index
ASCC	African Statistical Coordination Committee
CAPI	Computer Aided Personal Interviewing
CIS	Community Information System
CPI	Consumer Price Index:
CSEntry	Census and Survey Entry
CSPro	Census and Survey Processing
DHS	Demographic and Health Survey
EA	Enumeration Area
EPHI	Ethiopian Public Health Institute
ESA	Ethiopian Statistical Association
FA	Farmers' Association
FASDEV	Forum for African Statistical Development
FTP	File Transfer Protocol
GDP	Gross Domestic Product
GIS	Geographical Information System
GPS	Global Positioning System
HEWs	Health Extension Workers
IFPRI-ESSP	International Food Policy Research, Ethiopia Strategy Support Program
MDC	Mobile Data Collection
MDG	Millennium Development Goal
MoH	Ministry of Health
MTBDC	Mobile Technology Based Data Collection
PDA	Personal Digital Assistance
PHC	Population and Housing Census
PP	Producer Price
RP	Retail Price
SOP	Standard Operating Procedures
UNECA	United Nations Economic Commission for Africa
UoM	Unit of Measurement

1. Introduction

1.1. Background of the Study

In developing countries, where capacity is limited, surveys are often the only way to collect reliable data. Mobile Data Collection (MDC) is the use of mobile phones, tablets or PDAs for collecting data in real-time. There are many mobile applications (referred to as *platforms*) which allows to build a mobile data collection survey. These platforms allow customizing surveys to collect specific data as required, such as photographs, information from a list selection, voice recordings, GPS coordinates, and so on. Platforms vary in ease of use, cost, and features. Mobile technology can be used to employ long and complicated household surveys (Seebregts et al. 2009).

The Ethiopian Statistical Association, in collaboration with the Central Statistical Agency and the African Center for Statistics, UNECA has undertaken a pilot project on the use of mobile devices for data collection. The objective of the project was to increase the capacity of the Central Statistical Agency of Ethiopia in applying mobile devices for data collection so as to enhance the availability and access of timely and quality data in the country, to make exploratory investigation into the utilization of mobile technology for field data collection, and to strengthen working relationship between Central Statistical Agency and the Ethiopian Statistical Association, as a Training and Research Institute, in statistical development.

Statistical offices in Africa have been conducting surveys using paper-based data collection systems. Mobile devices are advantageous over the traditional manual enumeration with respect to the production of timely and quality data. Although mobile devices require using software tools that require collaboration with other partners they have the benefit over paper questionnaires as they can be easily tested after deploying the mobile-devices in the field (Arocena et al, 2013; Thakkar et al, 2013). Hence, implying that as the technology and innovation grows, national statistics offices and other data producers and users may have the capacities and skills to produce quality and timely data, produce and customize software packages appropriate for their National Statistics System.

Prior studies on testing the use of mobile devices for data collection confirm that the technology has tremendous merits with some little challenges. For instance, the initial one-time cost of mobile or tablet devices seems expensive and additional costs for maintenance and replacement due to

loss of devices can be acquired initially. But the devices can be used for future data collection season with no or very small cost. Set-up cost and ongoing cost, data security, data sharing and access and other related matters must be considered ahead of time. Studies also report that use of mobile devices simplifies monitoring quality and progresses of projects and even real time data collection campaign by enabling real-time access to data. Smart-phones have built in Global Positioning System (GPS) functionalities – which gives prominent opportunity to capture geographic locations at a time of enumeration. It is also environmental friendly as printing survey tools are significantly reduced or avoided. Devices are relatively small in size that they can be easily deployed to remote areas, and can be linked with other multimedia components such as audio and video tools. Hence, there is a reduction of logistics deployment and management and, very importantly, non-text data can be integrated with text data in real time. With mobile devices and the integrated software, data collection will be enabled with real time data check for data quality controlling using tools such as logical patterns, skip features and validation checks for answers entered (Arocena et al, 2013; Thakkar et al, 2013).

The newly emerged data types that are integrated into mobile devices include:

1. Geographic data: location, paths and boundaries;
2. Multimedia: photos, audio, recordings, videos etc.
3. Electronic sensors: finger prints, scanners, health sensors, smart cards and readers.

With these three data types, one can abstract additional indicators using GIS software to enhance data analysis. Locations and verifications functionalities enable the mapping of location, path, area or boundary to a geographical region. Prevalence and density functionalities in mobile devices facilitate the identification of certain activities or entities within a geographical boundary, for example the number of pharmacies in a slum neighborhood; forest-cover (density) per square kilometer (Matthew A. North 2011; Thakkar et al, 2013).

1.2. Merits of Mobile Data Collection

Paper-based data collection has been prone to errors. Storage costs, double data entry and reconciliation costs, and time elapsed for double entry are some of the challenges that have been compromising the production of timely, reliable and quality data. Gathering data from remote locations using paper-based collection methods can be very time consuming and it takes long time

for the data to reach users and analysts. Collecting accurate data is essential to get a true picture for resource planning and research. With the growth in availability and decline in cost of smart mobile phones, these seem an ideal technology to help to improve data accuracy and vastly reduce the time to reach those who need it. Cleaned data may arrive after the uses of the data have expired or after their importance is out-of-date for policy formulation and evidence-based decision making.

Thus, collecting data using mobile devices is ideal to enhance data accuracy, increase efficiency and reliability of the data and it can vastly reduce the time wastage and facilitates delivery of timely data.

The fact that smart phones are portable, have internet access, and can run third party applications makes them a natural fit for data collection and transfer. By exploiting internet capabilities of smart phones, real-time transfer of data collected using electronic forms can be achieved from remote areas to a center (Mosa et al., 2012). This might reduce the costs related to data processing, such as duplicating paper forms, carrying and storing paper forms, and data entry. Some of the advantages of mobile data collection over paper-based data collection are:

Speed: Mobile data collection speeds up the data collection process. Where network availability allows for a quick transmission of data to a server, the reduced amount of time that elapses between local data collection and delivery can save weeks or even months in the overall data collection process.

Accuracy: Data collected at the source can reduce transcription errors, and data transmission over mobile networks may ensure that no data are lost 'in transit'. Capture and transmission of data digitally may also ensure that it is easier to store and access them at later dates.

Convenience: Mobile phones as data collection tools offer convenience through their usability, size, and weight and battery longevity. Depending on the model used, they also offer additional features such as GPS location data and imaging

Training: Since the smart phone technology is growing sharply, users of the fantasy gets increased. This makes training users of smart mobiles about the device will be very easy. Self-reporting can also be employed using respondents' mobile device – which also reduces unnecessary resource wastage through buying mobile devices for surveys.

Power: Compared with devices such as laptops, mobile phones may be much easier to keep charged, as they require much less power and because many fast, low-cost charging options may be available in local communities because people are already utilizing such devices extensively in their daily livelihood.

Combining with other data: As a way of substantiating the information provided by text, data captured via mobile phones can be combined with data in other formats such as photographic images, audio, and video. In addition, GPS or geo-location data can be collected and transmitted along with survey data.

Low cost: as seen stated above, data collection through the use of mobile phones can be done at substantially lower cost than through traditional way. In addition to improving the efficiency of data collection efforts when compared with traditional, mobile data collection may also offer options for data collection that are simply feasible, or even possible, using other tools or methods.

Multiple Language Support: In diverse populations where multiple languages are spoken, multilingual paper assessments are cumbersome and costly. However, mobile phones easily integrate multiple language assessments with a simple drop down menu of language options.

1.3. Expected Challenges of Mobile Data Collection

Mobile data collection may have challenges or difficulties. Some common challenges include:

Device Security: Most mobile devices are so smart and costly which makes devices target for theft. Keeping the devices functional for a long period of time as an asset of an organization requires special care and device security policy. Device depreciation such as battery life, and device crash and loss of equipment can be risks that challenges application of mobile device for data collection unless special care and appropriate device management policy is set. Putting penalties into the contracts of surveyors and monitors is a common way to ensure equipment is well-looked after.

Data security: Data security and privacy are risks and challenges since data can be stolen or accessed improperly and the challenge may also risk data transmission process. The use of encryption at both the device level and during transmission can mitigate such risks. Data security tools, protocols, and good practices, as well as related regulatory frameworks should be rigorously

identified to overcome and avoid the risk before it happens. However, laws and guidelines governing the use of such tools, may not be widely known. Where tools are utilized during the mobile data collection process, care needs to be taken to ensure that ownership, possession and utilization of data collected and transmitted are clearly articulated. Sufficient mechanisms need to be in place to audit related arrangements and agreements, and to ensure that penalties for non-compliance are clear and enforceable.

Connection: Network connection and internet access are other expected challenges that matter data transmission and would likely impact MDC as it delays the data synchronization and data entry form online update.

Electricity and others: In most cases, fully charged devices batteries may not stay longer even for a day or half, based on the type and magnitude of the survey. Hence, batteries may require to be charged frequently. Thus, limited access to electricity at the vicinity of the EA affects the field work. Moreover, the limitation at the server configured for aggregating data significantly impacts the data synchronization, form update processes (whenever required) and data aggregation.

1.4. Considerations in Implementing Mobile Technology

1. Identifying right technology service provider
    - ✓ Hardware providers: Phones, computers
    - ✓ Software vendors for data collection, and if required for data collation and reporting
    - ✓ Data transmission service providers: Mobile internet
    - ✓ Web-hosting services for data aggregation and storage
2. Estimating costs
    - ✓ Hardware, software, transmission, aggregation, management costs
3. Planning timeliness
4. Training & piloting
5. Ensuring data quality
6. Ensuring data security

1.5. Objective of the Pilot Project

Given that mobile devices in data collection can improve efficiency in terms of cost and time as well as enhance the quality of data as compared to traditional methods, the main objective of this project is to increase the capability of the Central Statistical Agency of Ethiopia in applying mobile devices for data collection so as to enhance the availability and access of timely and quality data in the country. The project will investigate and evaluate the use of mobile technology for field data collection with the following specific objectives.

1. identify the right software technology
2. identify the right mobile device technology
3. develop procedures for receiving, collating and analyzing data
4. develop mechanisms for ensuring data security and quality
5. test data transmission and review field challenges and lessons observed
6. configure and test separate server for testing use of mobile devices for data collection
7. collate, merge and export data using other statistical software
8. keep database of the exported data at the server

2. African Experience about Mobile Data Collection

Some African countries have been applying mobile devices like mobile phones, PDAs, CAPI and other digital devices for data collection. For instance, a study in South Africa has successfully tested use of mobile phones enabling the devices for Java programming language to collect a health related household survey. This survey took place offline, uploaded data automatically by storing data secured until a network signal is accessed. The study also confirmed the possibility of incorporating multiple choice, free text, numeric, date, time and other question types (Tomlinson et al, 2009).

A review by the African Center for Statistics (ACS) and other studies reveal that several African countries have been using PDA, GPS and CAPI for different types of surveys and censuses such as Population and Housing Censuses, Agriculture and Livestock Censuses and Surveys, Census of Commercial and Industrial Businesses, digital cartography, different health related surveys, Youth Risk Behavioral Surveillance Survey, Price Statistics, employment, MDG baseline survey, food security, Feed the Future survey, Welfare monitoring surveys, Vital Statistics and other household surveys. Some of the countries that used mobile devices are Cape Verde, Senegal, Botswana, Burkina Faso, Cote d'Ivoire, Gabon, Morocco, Cameroon, Nigeria, Ethiopia, Togo, Benin, Kenya, Sierra Leone, Uganda, South Sudan, Burundi, Zambia, Somalia, Tanzania etc. (Thriemer et al 2012, Grameen foundation 2012, ACS 2014, PARIS21 2015).

The result of a study which compared PDAs versus paper based data collection in Tanzania witnessed that data entry via PDA was faster, cheaper by 25%, more accuracy of data, and no missing data was occurred with electronic data collection. The study further investigated that delayed data delivery and late error detections in the paper-based system have made error detection and correction difficult, while all these were avoided during the use of electronic data collection (Thriemer et al 2012).

However, such experiences have been implemented with some challenges. For instance, a review of implementations of different mobile data collection experiences in developing countries have reported several associated challenges such as network shortcomings, intensive software and hardware training for the users, collecting sensitive information as a matter of privacy, especially of patients, etc (Ganesan et al, 2012).

In addition to the above limitations discussed by Ganesan, PDAs are known to have shortcomings in terms of gathering qualitative data, in which, comments and narratives cannot be well taken as the nature of the devices limits the action. The other devices used in most of the studies and surveys indicated above were not smart phones with android applications, which implies that it doesn't include collection of multimedia and non-text data like pictures and GPS location units. Smart phones are assumed to constitute all these and other utilities enabling to deliver almost all in one application and services.

3. Ethiopian Experience about Mobile Data Collection

An assessment made for the purpose of this Ethiopia pilot study also shows the existence of local experiences in Ethiopia, and it has been observed that several institutions have been using smartphones for data collection and project evaluation purposes. WaterAid-Ethiopia, IFPRI-ESSP, EPHI, and Agricultural Transformation Agency are some on the list. Some of these institutions have used ODK software to develop data entry forms.

A study of maternal health care in Ethiopia, reported that health extension workers and midwives have found the electronic forms developed using ODK software on smartphones useful for their day-to-day maternal health care services delivery. However, the study has indicated the need for sustainable use and implementation of the technology with the required technical support to health workers and by securing mobile network airtime (Little et al 2013, Medhanyie et al 2015).

Even though, the above practices of using mobile devices for mobile data collection are growing in the continent, it is important to advance the system so as to enhance the quality and timing of official data production in the National Statistical Offices in Africa. Hence, testing the technology; developing or adopting compatible software; identifying challenges, risks and possible solutions or coping mechanisms; investigating data security and privacy issues, time and cost effectiveness; and observing lessons has enormous benefit for further application of advanced smart mobiles based data collection.

4. Project Design and Methodology

4.1. Concepts and Definitions

A mobile device is a handheld tablet or portable device both compact and lightweight. New data storage, processing and display technologies have allowed these small devices to do nearly anything that had previously been traditionally done with larger personal computers. Mobile devices ranges from 'low end phones' to 'Point of Sale' (PoS) terminals, each used to enter data and can be linked to add-on devices: biometric sensors, recording devices for finger print scanners, videos, pictures, smart cards.

Samsung, Sony, HTC, Motorola, LG and Apple are just a few examples of the many manufacturers that produce these types of devices.

Data collection software: Controls data entry based on programed formats and rules (repeating questions, skips, answer limits & validations, pre-loading of data). It can be installed for Android mobiles/ Windows tablets, and is not required in 'Low-end phones', which sends data through SMS or IVRS

Android is a mobile operating system. It is used in several smart phones. Unlike the iPhone OS, Android is open source app that developers can modify and customize the OS for each phone.

A Tablet is a wireless, portable personal computer with a touch screen interface. The tablet form factor is typically smaller than a notebook computer but larger than a smart phone. Types of tablets;

1. **Convertible tablets** have a display that rotates 180 degrees and can be folded to close, screen up, over an integrated hardware keyboard. Convertible models may allow user input through a variety of methods in addition to the hardware keyboard, including natural handwriting with a stylus or digital pen and typing through a screen-based software keyboard.
2. **Hybrid tablets are** sometimes referred to as convertible or hybrid notebooks. A hybrid is like a regular notebook but has a removable display that functions independently as a slate.
3. **Rugged tablets are** a slate-like model that is designed to withstand rough handling and extreme conditions. Rugged tablets are usually covered in a thick protective shell and have shock-protected hard drives.

Software refers to organized information in the form of operating systems, utilities, programs, and applications that enable computers to work. It consists of carefully-organized instructions and code written by programmers in any of various special computer languages.

SIM Card is a smart card that stores data for GSM cellular telephone subscribers. Such data include user identity, location and phone number, network authorization data, personal security keys, contact lists and stored text messages. Security features include authentication and encryption to protect data and prevent spy.

Transmission: Within a device, the distances between different subunits are too short. Thus, it is normal practice to transfer data between subunits using a separate wire to carry each bit of data. There are multiple wires connecting each sub-unit and data is exchanged using a *parallel transfer* mode. This mode of operation results in minimal delays in transferring each word.

Aggregation is any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables such as age, profession, or income.

Mobile Data Collection is a method of data collection using mobile devices on which software and mobile forms can be uploaded on mobile devices.

Data Management is the process of controlling the data generated during a research project. It is also a general term that covers a broad range of data applications. It may refer to basic data management concepts or to specific technologies. Some notable applications include data design, data storage, and data security.

Navigation keys: are keys that are found in the right and left side of a screen as an arrow and they used to move back and forth pages of questionnaire.

Retail Price is the money cost to the final purchaser of a specified unit of sale. It is inclusive of all sales and excise taxes.

Agricultural Producer Price is the price received by the producer/farmer for the transaction carried out at the first point of sale for a clearly specified agricultural product at the farm gate or nearest market. In most cases this price may not correspond to the farm gate prices as a farmer might sell his/her products at places outside his/her farm. The agricultural products have to be the farmer's own products.

Enumeration Area is a locality that is at most as large as farmers' association areas; it consists of about 150-200 households.

Price Index is defined as a measure of the weighted aggregated change in retail prices paid by consumers for a given basket of goods and services. Price changes are measured by re-pricing the same basket of goods and services at regular intervals, and comparing aggregate costs with costs of the same basket in a selected base period.

Table 1: Hardware devices: Options and Features

Options	Low-end phones	Smart phones	Tablets	Notebooks
Screen	Small gray scale	Touch screen	Large touch screen	Large screen
Data Entry	Keypad	Touch keyboard	Touch keyboard	Keyboard
Make Calls	Yes	Yes	Depends: Needs SIM card slot (space)	No
Mobile Internet	No	Yes	Depends: Needs SIM card slot (space)	No
Internet (Wi-Fi or Cable)	No	Yes	Yes	Yes
Apps and data collection software	No	Yes	Yes	Yes
Track location (GPS)	No	Yes	Depends: Some models do not have GPS capabilities	No
Capture photo and media	No	Yes	Yes	Yes
Connect to external devices (Printers, Finger print scanners, etc)	No	Yes	Yes	Yes
Hot-sync data to other devices	No	Yes	Yes	Yes
Battery life	Full day	Half day	8 – 10hrs	4 – 5hrs

Table 2: Data Transmission: Options and Features

Options	Mobile Internet	Internet	Hot-Sync
Set up time	None	None	None
Set up cost	None	Low: Cable or modem	None
Ongoing cost	Fixed: Data plan	Fixed: Data plan	None

Receive data from device	Yes	Yes	Cable connect
Send data to device on request	Yes	Yes	Yes

Table 3: Components: Summary

Device	Software	Transmission	Aggregation
Low-end phones	None	SMS	Remote server
Smart Phone	Data Entry App	Mobile Internet	Remote server

Table 4: Advantages and disadvantages: Summary

Device	Advantage	Disadvantage
Low-end phones	- Can work with any mobile handset - No hardware cost - Low set up time - Immediate access to data - No additional costs for scale up	- High SMS cost - Possibility of SMS entry errors (specially for longer content) - Quality control difficult
Smart Phone with mobile apps	- Easy data input interface - Low set up time and immediate data access - Control over devices and cost - Collect location or photo data for verification - Better data quality	- Handset costs - Ongoing cost of subscribing to mobile data plans (GPRS/ 3G) - Higher possibility of handsets to be stolen (expensive)

4.2. Survey tools and Study areas

This survey focuses mainly on monthly retail and producer price statistics. The survey used the existing questionnaire developed by CSA to collect data for both retail and producer prices. A CSPro software android version, CSEntry, will be used to develop the forms to be installed on the mobile devices and the data collection activities were done on the field by using mobile devices.

The study areas for this project are market places in Tigray Region. There are 8 market places in the region that CSA currently collects retail price data. In addition to the 8 market places, two enumeration areas were selected out of the existing agricultural sample survey enumeration areas in the same region.

4.3. Methods of Data collection

The price data collection uses weighing balance and scale meter to measure nonstandard items in the market areas. The producers' data will be collected from farmers who sold their agricultural products. All items for retail and producer prices used by CSA are included in this study.

4.3.1. Software, Installation and Application

The data collection requires tools such as mobile devices, software and external add-ons such as sensors, Power Bank chargers, SIM Card. CSPro Android Version, called CSEntry application developed by the US Census Bureau will be used for the study, using Sumsung Tab3, Android tablet.

4.3.2. Data Transmission

Mobile networks allow the transmission of data collected through SMS, voice, mobile-Internet, etc. With certain devices, like PoS terminals, data is transferred physically by hot-syncing cables. Data transmission to web-servers from mobile devices requires subscription to mobile Internet or data plans. Depending on the data-collection software, constant Internet access should not be required to store the data. The data can be stored on the device itself and transmitted whenever connectivity is available. For this study, a server will be configured at the Central Statistical Agency. IT specialists and Programmers will be assigned to follow up the progress of data transmission, collating and aggregating. Access to the server will be password protected to ensure security.

4.3.3. Data Security

All survey data will be collected and stored in electronic device, given that the mobile application will be password protected, and the data collected on it will be saved in raw string format. While this data was not encrypted, it could not be interpreted without access to the server-based questionnaire-formats. At the end of the survey and after the transmission of all data is completed, data will be automatically deleted from the phones memory. Furthermore, the data on the server will be password protected.

References

Arocena et al. (2013). *Utilizing Smartphone Potential for GPS Embedded Data Gathering Tool.* Proceedings of the 2nd Philippine Geomatics Symposium (PhilGEOS) 2013: Geomatics for a Resilient Agriculture and Forestry. November 28-29, 2013 |University of the Philippines.

ACS (2014). *Concept note: Ad hoc expert group meeting on the use of mobile devices for data collection.* 8 – 9 December, 2014. Tunisia, Tunis.

Ganesan et al (2012). *A Review on Challenges in Implementing Mobile Phone Based Data Collection in Developing Countries.* Journal of Health Informatics in Developing Countries. Vol. 6 No. 1, 2012.

Grameen foundation (2012). *Mobile Technology for Community Health in Ghana. What It Is and What Grameen Foundation Has Learned So Far.* Second edition: September 2012.

Little et al (2013). *Meeting Community Health Worker Needs for Maternal Health Care Service Delivery Using Appropriate Mobile Technologies in Ethiopia.* PLoS ONE 8(10): e77563. doi:10.1371/journal.pone.0077563.

Matthew A. North (2011). *An Empirical Analysis of Mobile Smart Device Accuracy and Efficiency In GPS-Enabled Field Data Collection. Issues in Information Systems.* Volume XII, No.1, pp. 318-327, 2011.

Medhanyie et al (2015). *Health workers' experiences, barriers, preferences and motivating factors in using mHealth forms in Ethiopia.* Human Resources for Health 2015, 13:2.

Mosa AS, Yoo I, Sheets L. (2012). *A systematic review of health care applications for smart phones.* BMC Med Inform DecisMak; 12:67.

PARIS21 (2015). *Informing a Data Revolution. COUNTRY REPORT ON CAPE VERDE.*

Seebregts CJ, Zwarenstein MF, Mathews C, Fairall L, Flisher AJ, Seebregts C, Mukoma W, Klepp KI: *Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a PalmTM Pilot interviewing system.* Int J Med Inform 2009 Volume 78, Issue 11, Pages 721–731.

Thakkar et al (2013). *Mobile-based technology for monitoring & evaluation. Reference Guide for Project Managers, M & E Specialists, Researchers, Donors.*

Thriemer et al (2012). *Replacing paper data collection forms with electronic data entry in the field: findings from a study of community-acquired bloodstream infections in Pemba, Zanzibar. BMC Research Notes.*

Tomlinson et al (2009). *The use of mobile phones as a data collection tool: A report from a household survey in South Africa.* BMC Medical Informatics and Decision Making, 9:51 doi: 10.1186/1472-6947-9-51.