EVALUATION OF SEED YAM PRODUCTION DATA COLLECTION METHODS AND OPEN DATA KIT PROTOTYPE FOR SEED YAM TRACKING IN IBADAN, NIGERIA

TURRY ATIENO OUMA

MASTER OF SCIENCE

(Research Methods)

JOMO KENYATTA UNIVERSITY

OF AGRICULTURE AND TECHNOLOGY

2021

Evaluation of Seed Yam Production Data Collection Methods and Open Data Kit Prototype for Seed Yam Tracking in Ibadan, Nigeria

Turry Atieno Ouma

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Research Methods in the Jomo Kenyatta University of Agriculture and Technology

2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature: ……………………………………………………..Date: …………………

Turry Atieno Ouma

This thesis has been submitted for examination with our approval as the university supervisors.

Signature: ……………………………………………………..Date: …………………

Dr. Agnes Mumo Kavoo, PhD

JKUAT, Kenya

Signature: ……………………………………………………..Date: …………………

Dr. Cornelius Mbathi Wainaina, PhD

JKUAT, Kenya

DEDICATION

This thesis is dedicated to Almighty God for granting me the opportunity, favour, wisdom and strength I needed to complete this journey. I also dedicate this work to my family, and especially my husband Amos Adalla who has constantly encouraged and supported my efforts towards successfully completing this thesis. To my children Jasmine and Malcolm who have been affected in all ways by this quest. To my mother Terry Mcodimba who was instrumental in giving words of wisdom when I needed it. I also thank my supervisors (JKUAT and IITA) who have constantly and actively participated to ensure my success in this venture.

AKNOWLEDGEMENTS

The authors acknowledge the Jomo Kenyatta University of Agriculture and Technology (JKUAT). Special thanks go to Dr. Cornelius Wainaina and Dr. Agnes Kavoo for tirelessly going through revisions and going out of their way to assist into getting to a final copy. This research was supported through a fellowship grant provided by the International Institute of Tropical Agriculture (IITA) and the CGIAR Research Program on Roots, Tubers and Bananas. Special mention goes to Dr. Trushar Shah and Dr. Lava Kumar who offered me the opportunity to conduct my research in IITA Nairobi and Nigeria, and who were instrumental in ensuring that I got all the support I needed to complete this thesis. The participants who took part in the survey and offered valuable feedback are also acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ACRONYMS AND ABBREVIATIONS

App	Mobile application
Enketo	An open-source web application
GPS	Global Positioning System
IITA	International Institute of Tropical Agriculture
ICT	Information and Communication Technologies
J2ME	Java 2 Platform, Micro Edition
NGO	Non-governmental organization
ODK	Open Data Kit
Ona	Data aggregation platform built on ODK
PAPI	Paper and Pencil Interviewing or just "paper-based"
SMS	Short message service
SPSS	Statistical Package for the Social Sciences
XLSForms	Tool used to simplify the creation and authoring of ODK forms in Excel

ABSTRACT

The purpose of this research is to illustrate the importance of applying open-source technologies for mobile-based data collection for real-time tracking of seed yam (*Dioscorea rotundata and Dioscorea alata*) production. Seed yam data in Nigeria have been predominantly collected using field notebooks in the past, which were subsequently entered into a Microsoft Excel spreadsheet or Access database for analysis and reposition. Relevant areas of weakness within the current data systems include time delay in providing feedback and non-availability of data upon demand. As a result, this has hampered direct tracking of seeds at various stages of crop production and management. The main research objective was to develop a seed yam-tracking application using Open Data Kit (ODK) that addresses these data collection challenges. The specific objectives were to evaluate the challenges relative to the existing data collection systems and reporting methods in seed yam production; to evaluate the effectiveness of the prototype developed for collection, entry and capture of seed yam production data; and to improve aggregation of field data and real time visualization of data using mobile technologies. The study provided data collectors with a pre-installed seed yam tracking app for use on an Android device for data collection and management. The system architecture was based on the ODK framework and a custom design based on requirements and lessons learned from literature. The ODK (Aggregate) made data easily accessible in minimum time possible from the point of data collection. The seed yam tracking application was designed as a result and field tested. The results indicated that this method improved efficiency, speed, and convenience in data collection and visualization, showing that this type of system can be used in crop management and research. The ODK system proposed for seed yam tracking is an important and effective solution that should be directed towards solving the identified constraints. Implementation of the system will streamline the process of seed yam data enumeration and aggregation, reduce greatly the time taken to provide information to stakeholders, it will act as an easier, user friendly and efficient system to provide feedback that ultimately informs decision making in the crop sector.

CHAPTER ONE

INTRODUCTION

1.1 Background

With the onset and advancement of technology, there has been considerable changes in the way data are generated and processed. Survey researchers have found new avenues to employ these technologies to advance data collection methods. However, many parts of the world are yet to gain from these developments. In the agriculture sector, for example, and where interventions are dependent on availability of localized data, the traditional method of data collection has predominantly been manual and often paper-based for collecting information on-farm or from farmers through personal interviews (Paper and Pencil Interviewing (PAPI or just "paper-based").

Data collected through the manual methods are not standardized and therefore are difficult to process for analytical and data mining purposes (Mechael, 2009).

Additionally, there are many resource constraints associated with paper based data collection methods including poor research infrastructure, personnel, time, costs associated with printing as well as costs for data entry and cleaning (Hudson, 2006).

1.1.1 Mobile technology in agriculture

While much effort has been made to overcome agricultural development challenges, there is a need for easily accessible critical data inputs that measure initial conditions and monitor progress for effective policy making to occur. Because of recent technological innovations in agriculture, there is a great deal of interest in using mobile devices to address existing information gaps. A number of studies have proposed mobile technology as a key method for improving data collection in the developing world and even in remote regions of developed world (Anokwa *et al.*, 2009; Baumüller, 2012; Hoogeveen *et al.*, 2014).

The affordability and availability of mobile phones and wireless networks make them a viable alternative to traditional paper and pencil methods. Through mobile technology, research surveys can be quickly deployed either by short message service (SMS) or mobile applications. In addition, responses can be visualized in a centralized place that may be miles away from the data collection site and pre-data analysis can begin while enumerators are still in the field. This ensures that less time is spent conducting data analysis (Fitzgerald and FitzGibbon, 2014). Results are therefore available faster, resulting in data-driven decision making by all stakeholders.

Compared with old technology, survey design (through programming and built-in data quality checks) based on modern, dynamic technology is more efficient and data collected are more refined, which saves time that would otherwise be used for data cleaning. This allows decision-making to have a more immediate impact. The switch from paper-based methods would streamline the process of data enumeration and aggregation, and provide immense potential for stakeholders with respect to reduction of error, improved efficiency and cost savings. This would help realize the full power of data and make sharing of information easier, thereby making the process attractive to individuals and organizations, including researchers, governments, non-governmental organization (NGOs), donors, and standard and regulatory bodies.

1.1.2 The yam food sector

Yam has become an increasingly preferred component of the Nigerian diet and many parts of West Africa Obayelu *et al.,* (2009). Nigeria is the largest yam producer in the world, accounting for >65% (38 million metric tons) of the world yam production (Verter and Bečvářová, 2015). Demand is, however, not met because of inadequate production and losses in storage. The predominant problems that plague the yam food sector hamper national policy program efforts geared at endorsing yam as a priority crop in Nigeria and in extension, West Africa. Increasing levels of on-farm losses of tubers during harvesting and storage, and pest and disease infestations are serious production problems. High cost of certified seed yam is however considered the most critical problem for farmers (Aighewi *et al.*, 2015).

Data on seed yam costs, pests and disease surveillance, availability of quality certified materials is not readily available in Nigeria and this has hampered a lot of efforts to improve seed yam production in the country Ironkwe *et al.*, (2008). One of the recommendations made by Sahel Capital (2019) to improve yam production was to strengthen the capacity and efficiency of regulatory agencies to support registration, and track and sanction mislabeling. Another recommendation was to enhance the capacity of informal processors and link them with markets. The difficulty in making concrete conclusions due to the lack/non-availability of official data on fresh yam and processed yam export was also cited.

1.1.3 Open Data Kit application

ODK is a modular, extensible, and open-source suite of tools designed to empower users to build information services for developing regions. ODK currently consists of four tools: Collect, Aggregate, Voice, and Build. ODK is designed to let users own, visualize, and share data without the difficulties of setting up and maintaining servers (Hartung *et al.*, 2010).

To increase yam production and improve the quality of yam produced in Nigeria, seed yam data should be easily collected and visualized in the minimum time possible. A switch from paper-based methods to mobile based methods has the potential to streamline the process of data enumeration and aggregation, and provide immense potential for stakeholders with respect to reduction of error, improved efficiency and cost savings. Mobile data collection methods, ODK in particular, if introduced in this context can help to ensure that there is availability of quality and validated seed yam information for decision making.

1.2 Statement of the Problem

Seed yam data in Nigeria has in the past been predominantly collected using field notebooks, subsequently entered into an Ms Access sheet for analysis and also reposition. Data collected through these manual methods are not standardized and therefore are difficult to process for analytical and data mining purposes (Mechael, 2009).

Relevant areas of weakness within the current data systems have been identified in the past. This include time delay in providing feedback to farmers, non-availability of data on demand, lack of validation of data, lack of multimedia capture (audio, video) and non-availability of real time location information (GPS). This hampers the direct tracking of seeds at various stages of yam production from field registration to, establishment, inspection, and storage of harvested yam. This also adds to the problem of seed yam production data not being readily available (FAOSTAT, 2014), meaning that it is hard to access information such as seed yam quantity, location, variety, geography, and availability. There is also a lack of a system that organizes seed production information, making it easier for institutions to monitor seed quality and certify producers. Additionally, there are many resource constraints associated with paper based data collection methods including personnel, time, costs associated with printing as well as costs for data entry and cleaning (Hudson, 2006)

1.3 Justification

Through the introduction of ODK software, this temporal delay between data collection and information output to farmers and seed regulators can be drastically reduced. Receiving useful feedback almost immediately is both an incentive for farmers to provide quality data and a rationale on which to base their management decisions to ultimately boost crop productivity. This feedback will also inform policy makers early and ensure immediate impact. One of the components of ODK (Aggregate) makes data easily available in the minimum time possible from the point of data collection.

Stakeholders can then visualize trends in the yam production as well as marketing and thus improve policy making and the general production of yam throughout the West African region.

1.4 Objectives

1.4.1 General objective

To assess the challenges of existing seed yam production data collection and reporting methods and to customize an effective ODK prototype for addressing the challenges in seed yam tracking

1.4.2 Specific objectives

1. Asses the challenges of existing seed yam production data collection and reporting methods

2. To evaluate the effectiveness of a customized ODK prototype in seed yam tracking

1.5 Research questions

1. What are the challenges with the existing data collection systems and reporting methods in seed yam production?
2. How effective would the prototype developed be in regards to data collection, entry and recording of seed yam production fields?

1.6 Expected outputs

The purpose of this thesis is to provide a customized ODK template that would help ensure efficiency in data collection, analysis and access; which would ultimately enable faster decision making, minimise data entry errors, reduce costs while supporting the seed yam certification process. The easy and effective system would inform, in real time, the status of seed production fields, compliance to standards, monitor deviations and understand bottlenecks for quality seed yam production in Ibadan, Nigeria.

Information collected would also facilitate field inspection and certification by the designated authorities. Information provided would ultimately be stored in the project database that can be availed by the project members.

The proposed ODK template would ensure that the data collected is cleaner by enforcing certain aspects of data quality. This includes edit checks programmed into the software which would ensure data meets certain required formats and ranges before the data is accepted into the database. This software would also ensure efficiency in data collection, analysis and access; which would enable faster decision making, reduce costs and can increase the efficiency and standards of the certification process. This model would then be adopted and extended to the certification process of the other crop species.

The proposed project would provide data collectors with a pre-installed ODK Collect app in two possible hardware platforms; tablet laptops with touch screens or keypads; or tablet mobile devices (Android tablet). This application would be designed using XLSForms; be able to download blank forms (designs) from an ODK Aggregate server, and upload completed forms (data) to the Aggregate server as well. By extension, it would allow the data collector to collect a number of data types including text, numeric data, GPS, photo, video, barcodes, and audio uploads to an online server.

CHAPTER TWO

LITERATURE REVIEW

The relationship between a research project and its results can be established only to the extent that relevant data are available. Methods used to collect data must be selected on the basis of the nature of the data required and the sources available as the utility of any research depends on its quality and purpose (Closs and Cheater, 1999). It is therefore important that a particular research question is matched with an appropriate design, the indicators used to capture the program's results and the type of analysis to be conducted.

A number of researchers have put forth literature supporting the transition from paper based interviews to mobile data collection methods by citing benefits and difficulties of the method. (Fitzgerald and FitzGibbon, 2014) for example, examined the potential benefits and challenges that may be encountered in transitioning from paper-based surveys to mobile data collection using handheld devices. They explored the implementation of digital data-gathering for an extensive panel survey in two different locations, one in Malawi, the other in Ethiopia. They conclude that the potential use of ICT in data collection in future will be pegged not only on the increased efficiency and accuracy, but also the reduction in costs from the use of affordable mobile devices.

Hudson(2006) assessed the general impact of the implementation and use of telecommunications within resource-constrained environments in four main classifications: ‘Efficiency’, ratio of output to cost (e.g., gathering info on soil/weather to improve agriculture yields); ‘Effectiveness’, increased quality of products, services, and organizational functions; ‘Equity’, distribution of development benefits to all areas; and ‘Reach’, the ability to communicate regardless of time or geographic boundaries .

(Sahin and Yan, 2013) systematically examined related literature using mobile phones as a tool for data collection. Their review indicates that there are significant benefits of collecting data with mobile phones over paper based surveys, the major one being that real time data can be received from actual settings by eliminating factors of memory. They however caution researchers on requiring participants to perform any activity that they would otherwise not perform because this can impede users' natural behavior and distort quality of data.

Vehovar *et al.*, (2010) explored two attributes of data collection: accuracy and efficiency of data collection, by examining costs and errors of implementing mobile phone surveys. They discovered that using mobile phone surveys was effective in reducing survey bias even though it was not effective in reducing costs.

From their evidence from the Tanzania and Sudan survey, (Hoogeveen *et al.*, 2012) assert that mobile based surveys can efficiently collect quality data that is useful to a wide range of data users including decision makers, statisticians, program managers, and researchers. The Tanzania survey puts forth the significance of putting in place tools that avoid attrition and non-response at the initial stages of implementing the baseline.

Dillon, (2012) highlighted some of the mistakes and take homes from a research project that used mobile phones to collect data from households in rural Tanzania. He also discussed the applicability of data collection with mobile phone in other settings.

Johnson (2013) assesses how mobile phones may enable better access to participants and higher participation rates in qualitative inquiry in Africa. With the explosion of mobile phone usage there is potential for qualitative researchers to reduce the number of ethical and procedural challenges they face when collecting data in the field.

Kwok (2009) explores the types of data that can be collected via mobile phones (including GPS data, Images, videos, voice recordings, and barcodes) every time and everywhere.

Researchers contend that ICTs may assist to realize development objectives in their roles as corresponding tools that contribute to the effectiveness of outreach programs (McNamara, 2003; Donner, 2008).

Muthiah *et al.,* (2012) identified key challenges of collecting data with mobile phones in rural areas in developing countries. He summarizes the challenges as follows: unreliable mobile networks, difficulty of saving data, mishandling devices due to lack of proper training, existing work load of health workers, and sustaining data accuracy, security and privacy.

Another group of researchers looked into challenges beyond the design process. For example, Brewer *et al.,* (2006) examined several challenges of technology research in low income regions, including technical, environmental, and cultural challenges.

Despite several efforts in research to address the data collection challenges using mobile technology, there is still room for improvement. The introduction of updated technologies opens new opportunities of improvement.

2.1 A review of some of the existing open source mobile-based data collection tools

A number of open source projects exist that seek to improve the data collection process. These frameworks have a worldwide community of developers, users and reviewers that share source code and improve them.

Below is a review of some of these tools in the context of low resource settings, poor research infrastructure and increasing demands for large scale crop production surveys with multiple data types.

2.1.1 Nokia Data Gathering

Nokia Data Gathering (Nokia, 2021) runs on Nokia and Windows phone devices. This platform only supports Nokia handsets, limiting it to Nokia users only. The main feature found in this platform is the survey question editor that allows multiple data types. Moreover, it offers ability to trigger questions based on the responses. This software was first developed in Brazil for monitoring the spread of Dengue Fever by the Health Vigilance Foundation. Since this tool is only available for Nokia users, it would be limiting in the area of study, therefore ODK was selected over it.

2.1.2 Open Data Kit

Open Data Kit (ODK, 2021) is an integrated, extensible, and open-source set of tools aimed at empowering users to easily build information services for resource contrained regions. ODK is made up of four tools: Collect, Aggregate, Voice, and Build. ODK is designed to let users own, visualize, and share data without the difficulties of setting up and maintaining servers (ODK, 2016). A number of software developers have built on this platform to provide paid for services.

2.1.3 openXdata

openXdata (openXdata, 2007) is a data collection platform supported by a syndicate of universities across the globe. It supports data collection using low-end Java phones in remote areas with small budgets. OpenXdata is a free open source platform that provides end-to-end solution and can work on all mobile devices. OpenXdata has the ability to support a wide variety of data types. OpenXdata requires some knowledge on MySQL and basic web application knowledge for deploying. It also relies on ODK Collect for data collection. OpenXdata also does not have an active support community online to consult in case of any errors. For these reasons, ODK was preferred.

2.1.4 EpiCollect

EpiCollect is an open source software for mobile data collection. EpiCollect is supported by Android and iPhone smartphones unlike ODK that is supported only by Android. EpiCollect can gather different types of data types, including text, audio, video, images and GPS data. EpiCollect is not dependent on network connectivity and this makes it a favourite in low resource areas. (Aanensen et al, 2009). While this application can be widely adapted for simple data collection and viewing, it lacks the engine to process linked complex forms as can be achieved on ODK. ODK presents a more intuitive form suitable for easier programming. These features led to the choice of ODK over EpiCollect.

2.1.5 RapidSMS

RapidSMS (RapidSMS, 2016) is a free and open-source. The structure of the tool offers functionalities of capturing data through SMS and managing the data through a web interface. RapidSMS does not need network connectivity to work, supports all types of mobile phones, as long as the device can send and receive a text message. RapidSMS does not support complex multiple data types, has no multi-lingual ability, data validation capabilities. RapidSMS has been used in a number of studies including national surveillance, response monitoring and supply chain (Sean Blaschke *et al.*, 2009).

2.1.6 JavaRosa

JavaRosa (javarosa, 2020) is an open source data collection tool that is built on J2ME. It runs on both smartphones and low-end phones. JavaRosa uses GPRS20 protocol for transfer of data from mobile devices to the storage server. It is a product of the OpenRosa consortium. It is based on the XForms standard (w3c, 2021) which is the official W3C standard for data collection and interchange. In some developing countries, people with low income are still using low-cost Java-based devices, however with the emergence of smartphones; the Java-based mobile users have significantly dropped. J2ME on which JavaRosa is built on, is an ecosystem that does not facilitate the development and use of software. This led to preference of ODK over JavaRosa.

2.1.7 FrontlineSMS

FrontlineSMS (Frontline, 2014) is an open source SMS-based data collection platform. It offers collection of data through text messaging. Despite the software being free, the user has to pay for any SMS sent. FrontlineSMS software does not need network connectivity to work, and was purposed to offer standalone, cost effective SMS oriented services to governmental and non-governmental organizations. ODK is free and therefore was an easy choice over FrontlineSMS.

2.2 Motivation to propose ODK for seed yam tracking

In the context of low resource settings and despite the existence of the other open source mobile data collection tools, the will to propose ODK for seed yam tracking was based on a number of reasons:

First, SMS based technologies like RapidSMS and FrontlineSMS were considered. These are open source tools that are interoperable and can publish data to web based servers through SMS messaging. It was noted that these tools have some limitations in user interface flexibility and mostly support simple text information and short forms. They do not require any prior installation of special software and therefore do not require high hardware overhead thereby being very ideal for short studies that have fewer questions (Frances *et al.*, 2010)

Tools that can accommodate long studies and support multiple data types with or without online connectivity were also considered ranging from Nokia Data Gathering which specifically works on Nokia phones only, openXdata which requires some knowledge on MySQL and basic web application knowledge for deploying, relies on ODK Collect for data collection and does not have an active support community online. JavaRosa, built on J2ME, an ecosystem that does not facilitate the development and use of software.

The fragmented value chain of yam requires larger forms with more questions and complex multiple question types like GPS, barcodes, images and sometimes audio and video. In this sense only EpiCollect and ODK were found to be capable of managing a large volume of information with good data validation measures and reduction in data entry error as compared to SMS based technologies that may demand tedious coding practices (Frances *et al*., 2010)

Most of the more robust and mature smartphone survey suites are commercial and must be purchased on a per phone basis. Additionally, most of the other tools use ODK in some way (FormHub, EpiCollect and EpiSurveyor, DoForms, Group Complete, KoBo Toolbox, Survey CTO, ViewWorld, PhiCollect) while others are ODK compatible (JavaRosa, OpenXData, RapidSMS).

For this thesis only free platforms such as EpiCollect and ODK were considered.

EpiCollect can be widely adapted for simple data collection and viewing, but lacks the engine to process linked complex forms as can be achieved on ODK. For seed yam tracking, one of the main requirements is to be able to link and integrate forms across the yam production fragmented value chain (production planning, planting, crop management, harvest and inspection). The proprietary nature of EpiCollect and non-extensibility is also limiting and this led to the use of ODK.

In the context of low resource settings and despite the existence of the other open source mobile data collection tools, the proposition of ODK for seed yam tracking was based on the following expectations: Affordability (free and open source), support complex questions with a variety of data types including binary, integer, and string, multi-lingual ability (to eliminate language barrier in the low resource settings), data validation capabilities, Online/Offline functionality, Android and web form enablement (interoperability), support for multiple question types (Image, Audio, Video, Scanner, and GPS data types for easier seed yam tracking) and active support community online.

2.2.1 Open Data Kit related work

The Open Data Kit was developed by the Department of Computer Science and Engineering at the University of Washington. ODK is a free and open-source collection of tools used for gathering and managing data in low-resource environments. A number of researchers have presented case studies in which ODK has several application areas including enterprises, community level solutions, and government, business, health and education sectors. Piette *et al*., (2011), assert that the integration of mobile technologies and open source frameworks to the health sector can tremendously improve health service. In another research Rajput et al., (2011) evaluated the efficiencies of adopting the use of ODK in data collection in resource limited settings and in particular, the health sector. They reported the application of ODK to be more cost effective compared to paper based methods.

Researchers of the Columbia University mechanical engineering department in New York City examined the use of smartphone software technology for questionnaire studies in international development. They surveyed 300 farmers in rural Mali using ODK equipped phones. They found ODK to be a convenient and helpful tool in the implementation of questionnaire based studies. They assert that the software has the potential to make a significant impact on the process of survey taking for development applications due to its exceptional versatility and free and open platform nature. (Frances *et al*., 2010)

2.3 Seed yam data

One of the emerging practices in agriculture is the integration of technology to measure or monitor field and crop conditions to enable farmers make informed decisions in various stages of the farming process and crop value chains. This in turn informs and enables stakeholders such as fertiliser manufacturers, extension workers, seed salesmen and crop consultants to examine this data easily and counsel farmers on optimising their farming practices to produce better yield and minimize wastage.

Data on seed yam production are not readily or easily accessible. However, according to FAOSTAT (2014) the quantity of seed yam required to plant fields in Nigeria and Ghana, the largest producers in West Africa, is estimated at 7 to 10 million tons. Traditional methods for seed generation are still applied by small scale farmers of yam. This comprises “milking” or harvesting the yam tubers early and using the resulting seed yams for planting. This normally leads to tubers that are of poor quality and sometimes diseased.

Another method applied by small scale farmers includes using the yam head or other parts of the yam that can sprout to produce seeds. Seed yam are also inherited from parents, purchased or acquired as gifts. The seed yams are then reproduced over and over until productivity becomes minimal from continuous exposure to pests and diseases (Aighewi et al., 2014).

However, these methods pose a challenge in that they do not result in a large number of yams, and often, reduce the quality of seeds. These practices can consequently be quite expensive for farmers going into yam production for the first time, because the cost of seed yam is relatively high. A lot of literature cites the limited supply of affordable good quality planting material as responsible for low productivity of yam (Ironkwe et al., 2007; Udoh et al., 2008; Ogbona et al., 2011; Asumugha and Ogbona 2013).

With the onset of mobile technology, analysis can easily be done and farmers can get instant access to information enabling them to make smarter decisions about crop production. Data generation in agricultural production will create enormous prospects for research and genomics. The availability of this data, created under real conditions, will allow researchers to continously and more proficiently optimize their prototypes, algorithms, and products.

2.4 Scope and Limitations of the Study

The general scope of this study was restricted to open source technologies, supported by an open-source community who contribute training documents, localization support, as well as additional tools. Due to the evolving nature of open source software development, the system's procedures and platforms and their functionality as described in this study may change over time.

The prototype was developed using the Android platform since this is the only platform used by ODK. The study was also limited to provide instructions based on the current functionality of ODK (as of Nov 2016) as opposed to a wide range of mobile data collection methods available in the world today.

CHAPTER THREE

MATERIALS AND METHODS

The study was conducted in Ibadan, Nigeria an area in which there is an increased yam research and extension work by the Institute for Tropical Agriculture (IITA) resulting in many seed yam producers adopting the minisett technique for the production of seed yams as a commercial enterprise.

3.1 Research design

The process of developing a prototype for seed yam data collection and reporting began by studying the open data kit framework, followed by designing custom functionalities for mobile data collection. System architecture for field data collection was designed on the basis of identified data accessibility, accuracy and efficiency needs of crop management organizations, producers and field personnel. 3.1.1 Architectural features of the prototype

The architecture used in this study had the following features:

(i).Free and open source

This was paramount in the resource setting where this study was conducted and in light of the fact that researchers in seed yam production depend on donor funding.

(ii).User-friendly

In the context where this study was conducted and where small scale seed producers were some of the stakeholders, it was important that the tool be user friend to ease adoption

(iii) Robust and interoperable

Since the stakeholders included seed regulators and researchers, it was important to have a tool that could be used both in an office set-up (web form) and in the field (using mobile devices)

(iv) Provides accurate data

The collected data would be used for decision making in the yam food sector and therefore it was important that the tool should have data validation checks to ensure errors were minimised and less time and money spent on cleaning the data.

(v) Possesses offline capabilities

Given the fact that the rural areas of Africa have limited resources and there is a high possibility of intermittent internet connectivity, it was necessary to have a tool that would be used offline

(vi) Usable for multiple data types

For seed yam tracking, it was important to have a tool that has the ability to handle multiple question types such as barcodes, audios and GPS which were paramount in linking forms as well as providing real time tracking of fields

(vii) Ability to capture geo-referenced data using global positioning system (GPS)

This was very useful for the seed regulators when locating seed producers and real time tracking of seed yam

(viii) Data security

Since the data collected is sensitive and contained personal information, it was important to have a system that would guarantee data security.

3.1.1 ODK tools and installation

The ODK tools, Build and XLSForm were used to design a survey form. The ODK Collect was installed to download and fill-in the survey using any Android mobile device and aggregation server was deployed for hosting the survey form and gathering the survey data.

3.1.1.1 Survey form design (XLSForm/XForm)

This involved definition of indicators and the corresponding questions in the questionnaire, formulation of variables and identification of data types and value labels. The exact functionality and display style of each question would also be specified in the XLSForm definition using the type and appearance columns (**Appendix C**).This would ensure that each question was validated and only received the right data type. It would also make the user interface friendly and enable data validation checks.

Programming the tool followed input of validation checks, linking of separate forms using barcodes and enforcement of "pull data" function to ensure that every form could provide previous collected data at a different stage of seed yam production at the click of a button. Customization of the ODK tool to reflect seed yam tracking was also done using APIs and images. Reports were also designed and enabled for easier sharing of information from the system.

3.1.1.2 Deployment of survey forms on ODK Collect Android App

Once the XLSForms were ready, they were deployed on the ODK server for data collection. The next step involved installation of ODK Collect on an Android-based mobile device.

ODK Collect was downloaded from Google Play Store and configured for use.

The designed forms were then downloaded to the mobile device, ready for data collection.

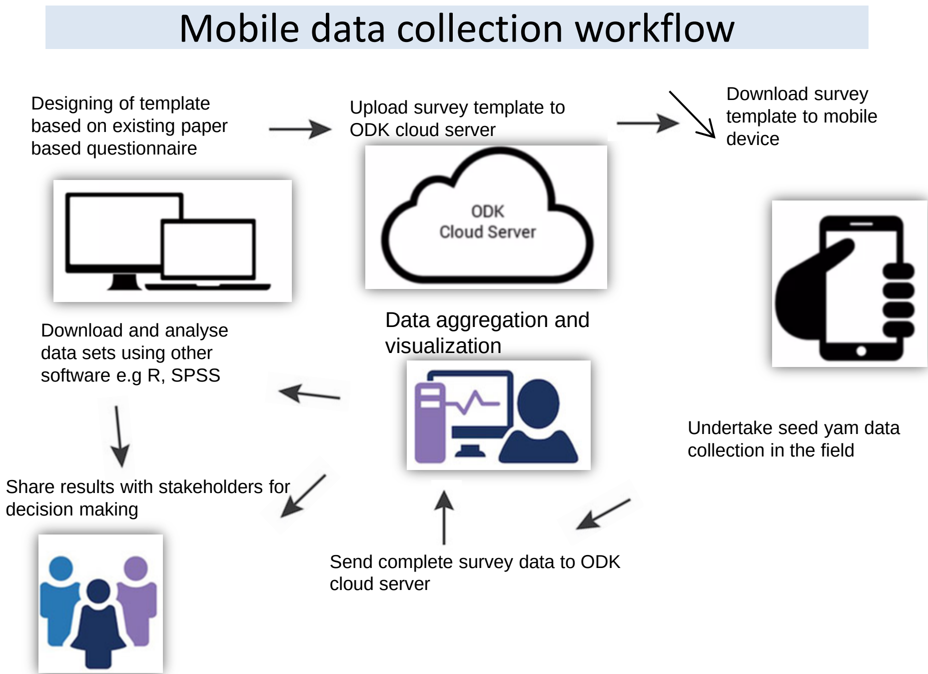

Figure 3.1: Mobile data collection workflow

3.1.1.3 Training of data collectors and data collection

Two researchers at IITA Ibadan, working on yam production were trained on the research survey questionnaire and the proper use and management of mobile devices as well as ODK Collect. The training involved the proper use of mobile devices for data collection, ethical issues to consider and management of mobile devices. These researchers were picked on the basis that they worked in yam production and were knowledgeable about seed yam tracking. This took a period of two days. The XLS Forms were deployed and the survey was hosted and survey results gathered and aggregated on an ODK compatible server.

3.2 Target population and sampling design

Literature search was conducted to identify stakeholder categories, identified stakeholders were represented in a visual chart, and correspondence established with expert informants to obtain practice-based insight. Key Informant Interviews (KII) with researchers and IITA staff allowed the identification and characterization of stakeholders of the seed yam sector (both formal and informal).

Following (Jiggins and Collins, 2003), stakeholders were grouped into three classes: primary, intermediary and key stakeholders. From among a list of the stakeholder categories, the following three potential categories of users of seed yam data emerged based on various stages of seed yam tracking and active and continuous use and need of seed yam data: 1) Producers, 2) Regulators of seed, and 3) Researchers.

Key informant interviews indicated that there were ongoing registration of seed producers and regulators at the time of the study. As such there was no sampling frame available for use, the purposive and snowball sampling techniques were employed to ensure representativeness of the users of seed yam data in the area of study. Potential users were identified through sensitization trainings, which were essentially to introduce the users to the prototype. These users were then tasked to suggest/recruit other potential users (producers, field regulators and/researchers). A total of 75 respondents were identified, i.e., 41 producers, 24 regulators of seed and 10 researchers. All (100%) of the respondents were from Nigeria.

3.2.1 Data Collection instrument(s)

Both primary and secondary data were utilised in the study. Primary data were generated through a well-structured questionnaire (**Appendix B**), which was administered both online (through the ODK Enketo web form and also using mobile devices with the ODK Collect app installed. The interviews were conducted between July and August 2018. Seed yam secondary data were sourced from literature and relevant research works in the area. This included previous research work conducted by IITA on seed yam production using new technologies. This provided a means of identifying the bottlenecks in seed yam tracking.

3.2.2 Pilot study

To better understand gaps in existing practice, this study reviewed existing methods used in field data collection. Initially, crop researchers at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, were interviewed to learn about the current practices used in data collection, type of information collected, and variables that were considered important. The paper-based questionnaire that had previously been used to collect seed yam data was presented, variables and data types identified and then used to design XLS Forms that would be applied for mobile data collection.

3.2.4 Needs assessment

To address the first objective of this research project, a needs assessment of the system and its applicability to crop field-data collection was made. The system was pilot tested via two data collection exercises with the stakeholders who had been identified as primary users of seed yam data (producers, seed regulators and researchers) but who were not among the sample selected. Feedback from the data collection team during this period was to be used to improve the system.

Upon reviewing the system, the sampled respondents were asked to fill out a short questionnaire (**Appendix B**). This included questions on predominant data collection methods used and challenges faced. It also had questions on what device they used, some questions to evaluate the effectiveness of an ODK prototype and whether or not it satisfied their requirements for a mobile data collection system. There were also questions on how to improve the system and a request for suggestions on what user interface features could be included in the future.

3.3 Data analysis and presentation

The results of user evaluation and assessments were analyzed and used as a basis for determining the usability of the application. Descriptive statistics (frequencies, tables, means and percentages) were employed in the analysis. Quantitative analysis was conducted using SPSS software and data visualized through Microsoft Excel.

CHAPTER FOUR

RESULTS

4.1 Designing of XLSForm for mobile data collection

A total of 75 respondents were interviewed, as shown in Table 1. These primary stakeholders were directly affected, either positively or negatively by seed yam projects or interventions in the yam sector.

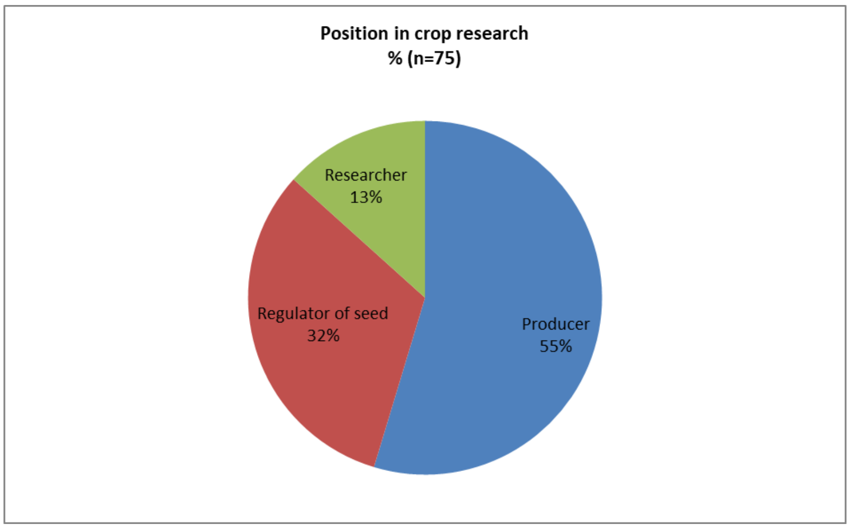

Figure 4.1: Categories of survey sample population

The XLSForms were designed (**Appendix C**) and the survey forms were deployed on ODK Collect Android App (Fig. 3) and could also be accessed via web forms (Fig. 4)

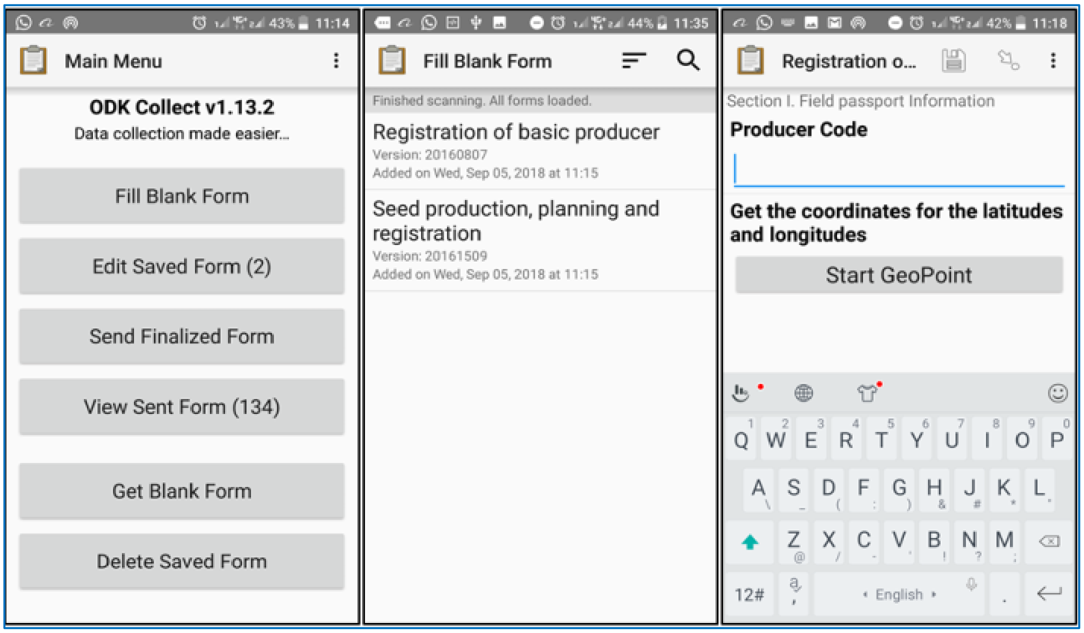

Figure 4.2: ODK Collect option

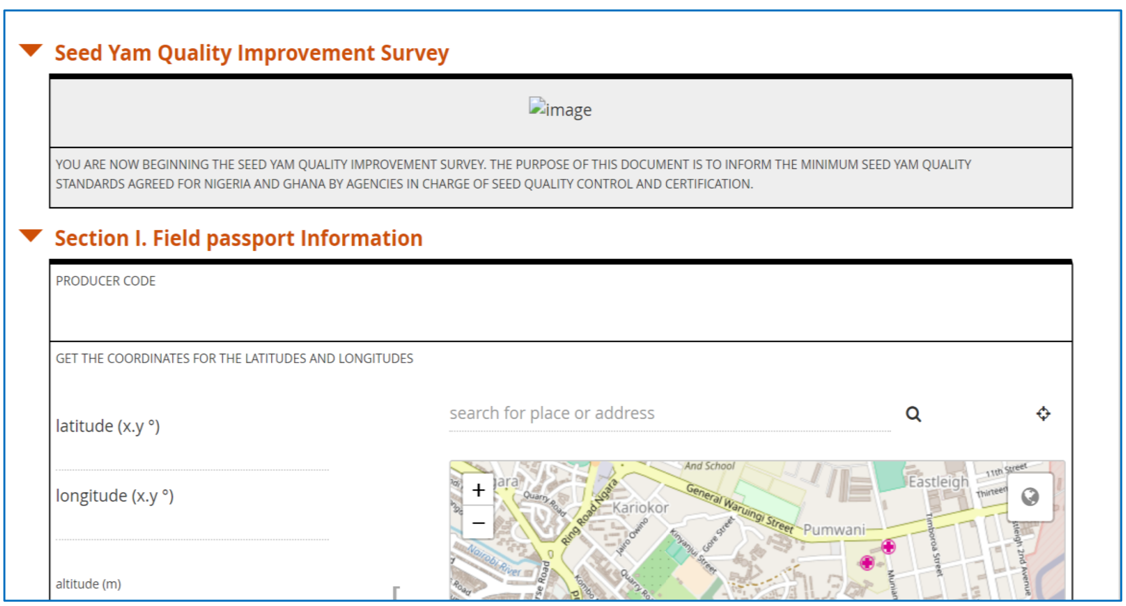

Figure 4.3: Enketo web interface for data entry and visualization

4.2 Evaluation of existing data collection systems and reporting methods in seed yam production.

4.2.1 Predominant data collection method

The most predominant method of data collection within the country was ‘paper method’, with 79% (59 out of 75) of the respondents citing this. A small proportion (4%) of the respondents mentioned that both paper and mobile means of data collection were used. When further probed, the respondents seemed to think mobile-based data collection was the use of SMS for collecting data. Researchers and regulators of seed reported to have been applying both paper and mobile based methods (Fig. 5)

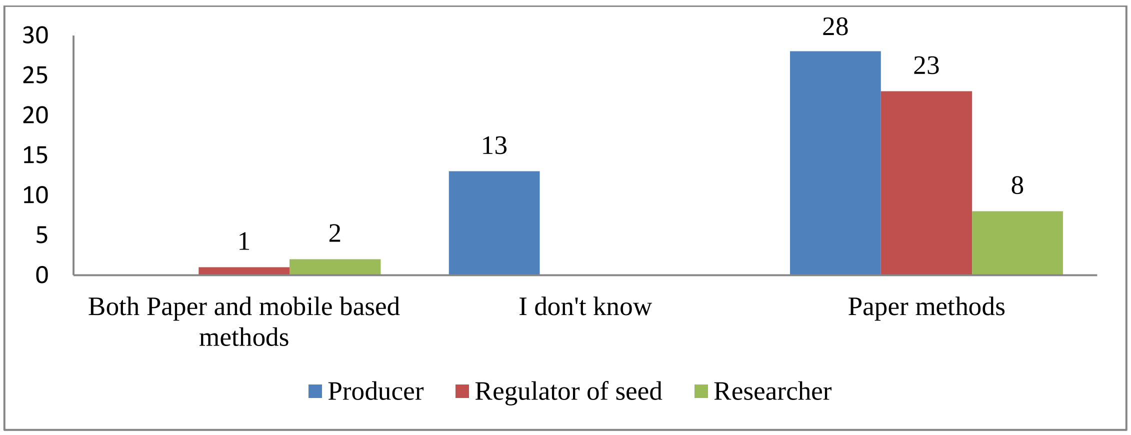

Figure 4.4: Predominant data collection methods disaggregated by users of seed yam

4.2.1.1 Efficiency of the predominant method of data collection

When asked about efficiency of the predominant method on a scale of 1-10, with 1 being very inefficient and 10 being very efficient, the respondents who had cited paper-based methods rated the method 5, which means the method was thought to be moderately efficient.

Respondents who had cited both paper- and mobile-based methods gave a rating of 8, which means that a combined use of both methods increased efficiency.

4.2.1.2 Feedback rating of the predominant method of data collection

Effective feedback, both positive and negative, is very helpful. Feedback provides valuable information that will be used to make important decisions. The time taken to receive feedback is one of the aspects that necessitated the introduction of mobile data collection. When asked how long it normally takes to receive feedback using the predominant method, a majority (53%) of the respondents indicated that it took between 3 to 6 months to receive feedback and sometimes even up to one year as shown below in Table 1.

Table 4.1: Time taken to receive feedback using the predominant method of data collection

Time taken to receive feedback	Respondents (%)
Between 3-6 months	40%
Between 6 months-1 year	35%

When the feedback data was disaggregated, it became evident that a majority (68%) of seed producers were the last to receive feedback and the delay would last up to a year. a majority of seed regulators (88%) and researchers (60%) were able to receive feedback between 3-6 months (Fig.6).

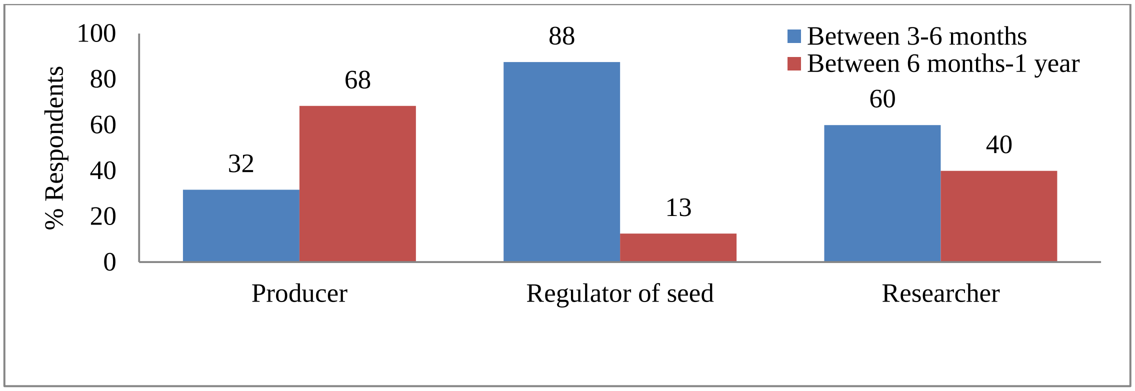

Figure 4.5: Time taken to get feedback, disaggregated per primary user

4.2.2 Use of Mobile Technology

Mobile technology is used to a large extent in the country, with 100% of the respondents affirming this. Smartphones are the most common mobile phones (61%) used according to the respondents (Fig. 7).

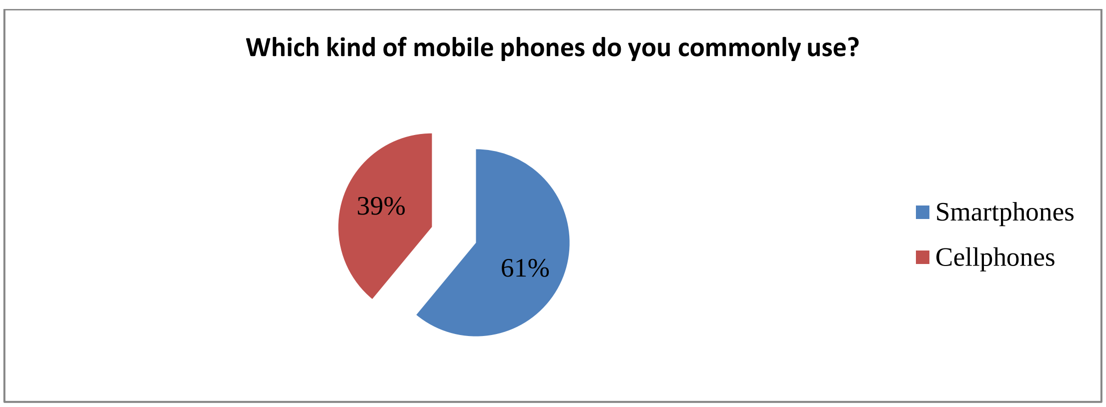

Figure 4.6: Categories of commonly used mobile phones

A smartphone's capability is more oriented towards mobile phone options and has much more than basic cell phone capabilities. Thirty-nine percent (39%) of the respondents mentioned that they used mobile phones. Among the 39% mobile phone users, 81% alluded that mobile device applications could make seed yam data collection and reporting more efficient (Fig. 8).

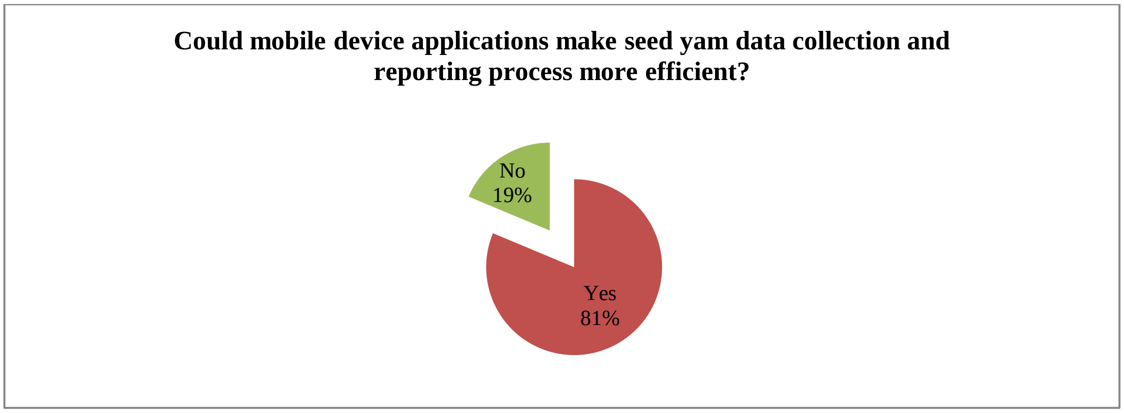

Figure 4.7: Could mobile device applications make seed yam data collection efficient?

Producers reported mixed answers, a number (34%) not being sure whether mobile device applications would make the reporting process efficient as seen in Table 2 below.

Table 4.2: Could mobile device applications make seed yam data collection efficient?

Response	Response by user category (%)			
	Producer	**Seed regulator**	**Researcher**	**Grand Total**
Yes	66	100	100	81
No	34	0	0	19

ODK uses the Android platform and therefore it was important to know whether the intended users were already using this platform. It was evident that this was the case, with 92% of the respondents affirming that they used the Android platform; 8% used Nokia/Microsoft Windows-based platform.

4.2.3 Challenges with the existing data collection methods

Results from this study revealed a number of challenges with the use of existing data collection methods. An open ended question was asked to explore the challenges of the existing data collection methods. Most of the crop information systems (53.3%) were not computerized. This would severely hamper the process from data collection to analysis. A large percentage of the respondents (67%) reported that the biggest challenge was unavailability of data (Fig. 9) and Table 3).

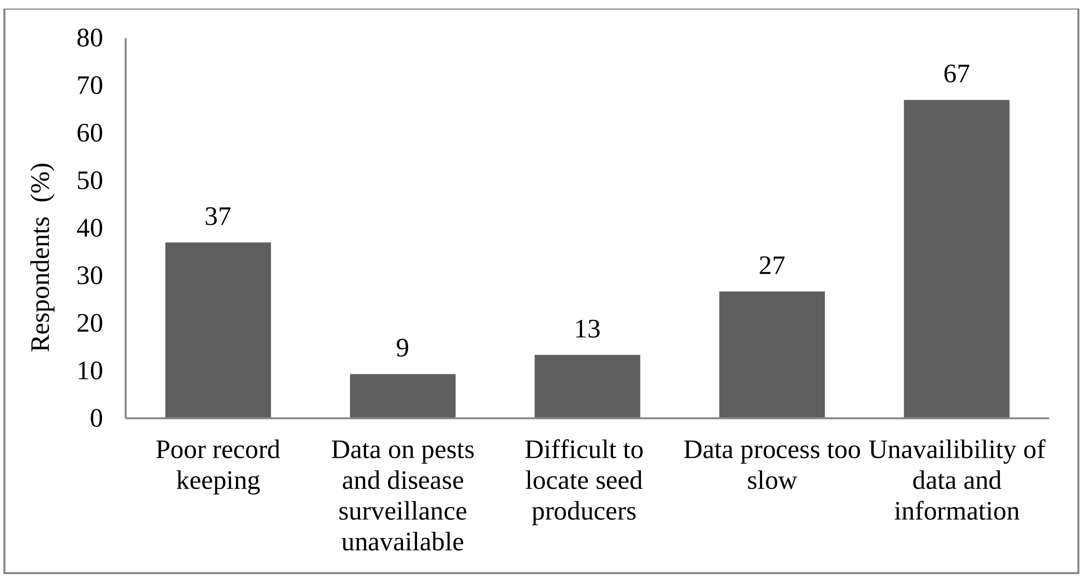

Figure 4.8: Challenges with the existing data collection methods

Disaggregation of data on he challenges with the existing data collection methods by seed yam user category revealed that crop information was not readily available when requested for. Other challenges included poor record keeping and a long time-lag between data collection and reporting. Seed producers in particular reported that they had no access to seed information such as seed availability and prices.

Table 4.3: Challenges with the existing data collection methods by user category

Challenge	**Respondents per user category (%)**		
	Producer	**Seed Regulator**	**Researcher**
Costs of data collection and analysis too high	0	2	8
Data processing too slow	15	42	40
Difficult to locate seed producers	15	13	10
Data on pests & disease surveillance unavailable	0	29	0
Poor record keeping	7	0	10
Feedback is not easy to get	2	0	0
Funding for research limited	0	4	20
Paper based systems are hard to manage	2	0	0
Unavailability of data and information	59	13	10
Total	100	100	100

100% of the respondents interviewed mentioned that it was very difficult to access seed yam data. On a scale of 1-10 (where 1 = very difficult and 10= very easy), mean score for all 75 respondents was 3.37 ± 1.67SD.

4.3 Evaluation of the effectiveness of the developed ODK prototype in seed yam tracking

4.3.1 Piloting of the ODK prototype

In order to measure the perceived effectiveness of the developed ODK prototype in seed yam tracking, it was important to pilot its usability among potential users. The users were shown the application and how to navigate through the pages. They were also allowed to freely explore the prototype.

However, the target population was initially interrogated on their expectations of the prototype. 75% percent cited user-friendliness and data accessibility in near real-time as their top expectations. Other expectations cited were the ability to collect GPS data and offline capabilities of the application. The rest of the respondents were concerned more with adaptability and SMS and/or email accessibility.

4.3.2 Evaluation of the proposed system/prototype

After the exploration of the proposed system, all respondents (100%) reported that the proposed application would be useful for addressing their crop production data needs. The respondents gave varied reasons for preference of the application (Table 4). The most outstanding reason was user-friendliness of the application and the fact that data would easily be accessible.

Table 4.4: Evaluation of the proposed system by various categories of seed yam users

Characteristic of application	User category respondents (n=75)			
	Producer	**Seed Regulator**	**Researcher**	**Total**
User friendly	20	3	4	27
It makes data readily available	5	6	12	23
It makes information very easily accessible	6	6	4	16
Minimizes time taken from collection to reporting	2	3	4	9
Offline capabilities	2	3	3	8
Supports multiple data types e.g maps, video, audio	1	3	3	7
Makes work easy	1	2	2	5
Immediate data visualization	1	1	1	3
Easy to get feedback through SMS and email	0	0	1	1
Total	38	27	34	99

The proposed system was also assessed on the basis of ease of use, presentation characteristics, information content and general design. The answers to the questions revealed that most of the participants found it both easy to learn and easy to use. This assessment was based on a 5-point Likert scale, ranging from strongly agrees to strongly disagree (Table 5 and Fig. 10)

Table 4.5: Rating of the prototype presentation characteristics (%)

Prototype presentation characteristics	Strongly Agree (%)	Agree (%)	Neither Agree nor Disagree (%)	Disagree (%)	Strongly disagree (%)
The prototype instructions are logical, clear, consistent and understandable for a layman	16	73	11	0	0
The prototype design, information content, and performance is up to par to collect and visualize results	24	73	3	0	0
The layout of text, graphics, figures, and links is user-friendly	52	47	1	0	0
The colours used in the prototype are appropriate	53	31	16	0	0
The text size and font are legible	49	47	4	0	0
The prototype provides an efficient and usable means of accessing, querying, analysing, and interpreting data	45	47	8	0	0
The prototype addresses the challenges of crop information systems	27	73	0	0	0

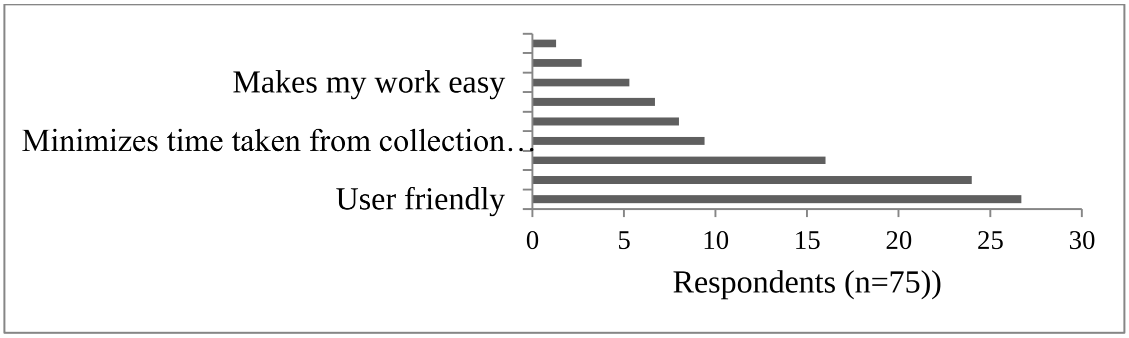

Figure 4.9: Evaluation of the proposed system (prototype)

In addition to the features that the proposed prototype offers, the respondents were asked to cite any other features that could be included in future to improve data analysis and reporting process.

Their suggestions, listed below, included both improvements to enable ease and speed of access of information, the user interface for user-friendliness and further food for thought on what is required for adoption of the system (Table 6 and Fig. 11).

Table 4.6: Suggested features to be added to prototype by various seed yam user categories

Suggested features	Respondents (%)		
	Producer	Seed regulator	Researcher
Feedback through SMS, alerts on price fluctuations	5		
It is ok the way it is	20	10	10
Involve the government in spanning out the application	5	6	0
Provide mobile phones to producers for a wider coverage	6	4	3
Should have offline capabilities	6	2	
Give incentives to seed producers for them to use the application	7		
Allow printing of output	4	6	2
Provide training/tutorials	2	2	0

The results also revealed that 40% of the respondents were satisfied with the prototype (reported that the prototype was OK) without any additional features and/or revisions while the rest suggested various features that would make the prototype more effective (Fig 11)

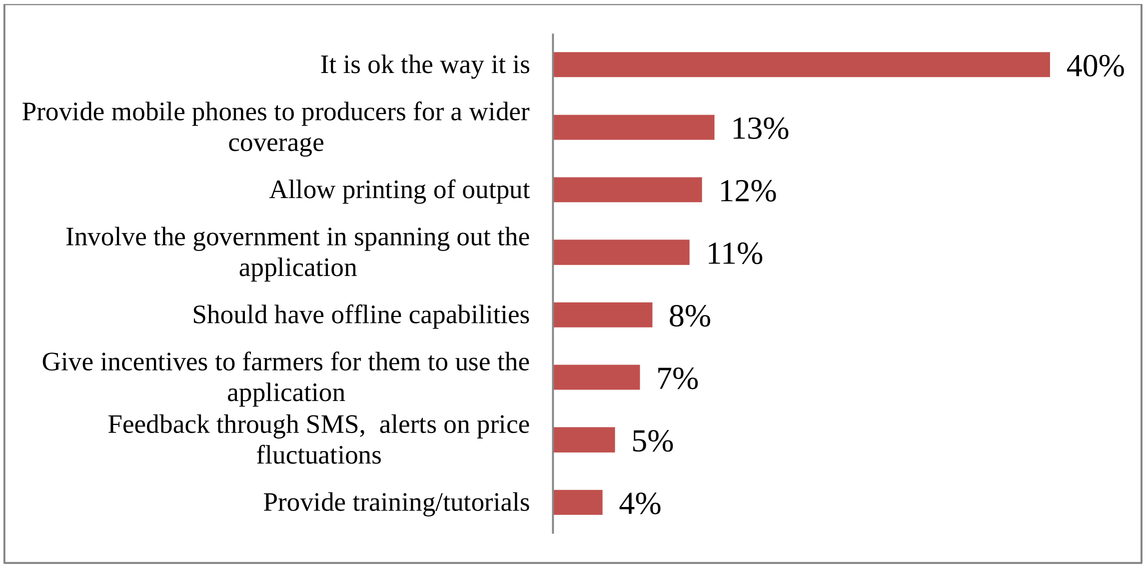

Figure 4.10: Respondents' suggestions for additional prototype features

4.3.3 Limitations of proposed system/prototype

The respondents were asked to give any limitation of the prototype design with respect to data access, analysis, and interpretation. Results are shown below in Table 7 and Fig. 12

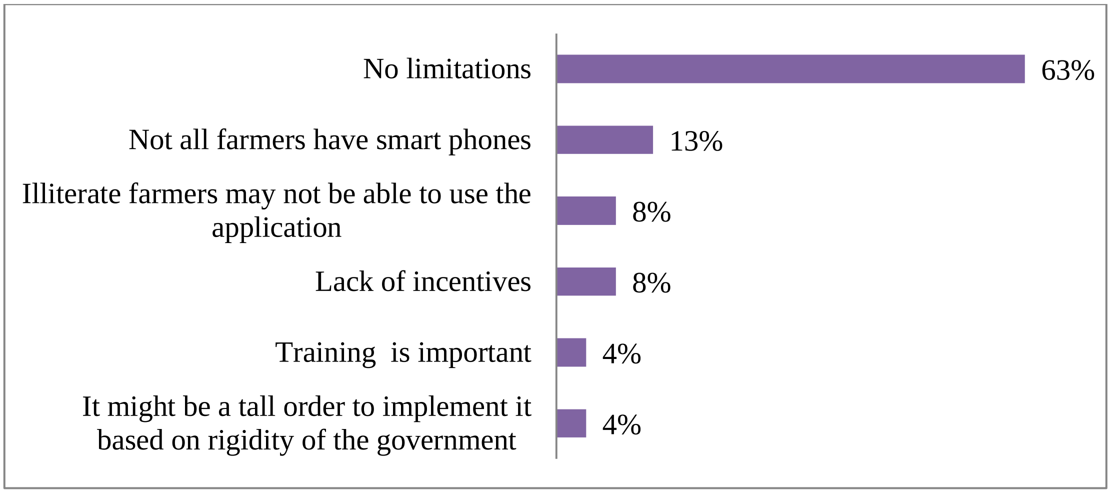

Figure 4.11: Perceived limitations of the proposed system

As opposed to the predominant data collection methods, the ease with which data was accessed increased tremendously as majority of the respondents rated the ease of access at between 7-10 on a scale of 1-10 with 1 being extremely difficult while 10 being extremely easy ((Fig. 13).

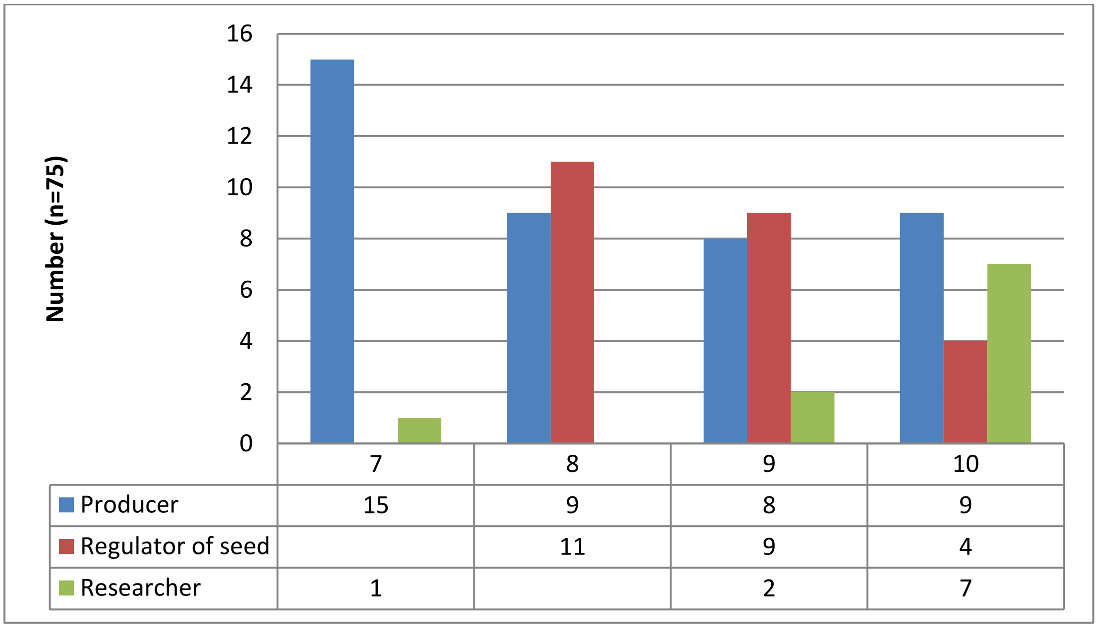

Table 4.7: Perceived limitations of the system by various categories of seed yam users

Perceived limitations	Respondents (%)		
	Producer	Seed regulator	Researcher
Lack of incentives	8		
Illiterate farmers may not be able to use the application	6	1	1
It might be a tall order to implement it based on rigidity of the government	3	1	
No limitations	41	15	7
Not all farmers have smart phones	8	4	1
Training is important		1	3

4.3.4 Improvement of aggregation of field data and real time visualization of data using mobile technologies

Aggregation and visualisation of data is one of the most impactful ways data analysts and scientists can communicate their findings. A data visualisation assessment was conducted to ascertain the prototype's capabilities to handle data visualization requirements. A 5 point Likert scale was applied ranging from Strongly Agree to Strongly Disagree (Table 8)

Table 4.8: Assessment of the prototype's data visualization characteristics

Data visualization characteristic	Strongly Agree %	Agree %	Neither Agree nor Disagree %	Disagree %	Strongly Disagree %
Data was presented in a logical and clear manner that was easy to interpret	37	52	11	0	0
Data was visualised immediately after submission as long as there was internet connectivity	52	37	11	0	0
Descriptive data was visualised in form of graphs, charts and tables	55	37	8	0	0
The prototype would be useful in decision making for stakeholders	88	12	0	0	0
The prototype facilitates the search, retrieval, sharing, analysis, and visualization of seed yam data	84	16	0	0	0
The prototype addresses the challenges of crop information systems	100	0	0	0	0

CHAPTER FIVE

DISCUSSION

5.1 Evaluation of existing data collection systems and reporting methods in seed yam production

The most predominant method of data collection was found to be paper based method. This is expected based on existing literature which indicates that the agricultural sector still employs the traditional data collection methods. A number of challenges were associated with this method including a time delay in receiving feedback with a majority of seed producers mentioning that sometimes it would take up to a year before getting feedback.

Data collected through paper based methods are generally not standardized or validated and require data entry and cleaning before analysis. There are also many resource constraints associated with this method that would cause this delay. These constraints identified by the stakeholders agree with the previous research findings by Hudson, 2006 who reported some of the limitations associated with paper based methods, including poor research infrastructure, personnel, time, costs associated with printing as well as costs for data entry and cleaning.

This is also in agreement with Mechael, 2009 who also argues that data collected though paper based methods are difficult to process for analytical and data mining purposes.

The respondents also cited unavailability of data and information as the biggest challenge posed by this method. This means that important information that should reach primary stakeholders of the yam sector such as seed yam quantity, location, variety, geography, and even availability of quality certified materials is not easily accessible. This agrees with (Mignouna et al., 2014) who cited the lack of market information as the biggest drawback for Nigerian agriculture. As a result, farmers are in a predicament as they are unable to attune their production practices to market changes.

Seed regulators mentioned that sometimes mapping of seed producers' locations in various locations in Ibadan because of lack of real time GPS data. They also mentioned that crop disease surveillance was hampered because it took longer to locate, collect and analyze data from different locations. This is expected as paper methods do not allow a variety of question types such as GPS, barcodes, videos or audios.

The interviewed respondents cited difficulties in accessing seed yam data. This supports the observation by Kenyon, L., 2009 on the scarcity of reliable data on the area, production and yield of yam.

5.2 Evaluation of the effectiveness of the proposed prototype in seed yam tracking

The respondents cited user-friendliness as one of their top expectations for an ODK prototype in seed yam tracking. This is in line with Hartung et al., 2010 who observed that to truly address the information gaps in developing regions, any new information service should be easy to use by minimally-trained users. Offline capabilities were another expectation cited among the top five perceived expectations. One of the concerns that researchers have when considering replacing paper questionnaires with digital tools is if the tool will work in remote areas where there is limited to no internet connectivity.

There were suggestions of improvements to the prototype including providing incentives and mobile phones to seed producers and involving the government in scaling the proposed ODK prototype. This is expected since mobile phone usage heavily relies on adaptive capacity. This is the capacity of individuals or groups to manage and influence their resources and risks in the face of a perturbation (Waters & Adger, 2017). A number of determinants of household-level adaptive capacity have been noted in research and they include access to assets and resources (Moser, 1998).

5.2.1 Designing of the prototype

In the design of the prototype, the goal was to address the challenges posed by paper based methods, meet the users' expectations while leveraging on the ODK platform to offer open-source and standards-based tools which are easy to try, use, modify and easy to scale. The first step in the design was specification of the data types. This included string, numeric, bar codes for easier identification & verification of certified seed, images to confirm placement of bar codes and vegetation around, videos to record any qualitative data and GPS to record location for easy identification of producers' farms.

To make this prototype widely accessible and functional, the design allowed the user to switch through multiple languages which in the case of Ibadan was Yoruba, Igbo, Hausa and English. The Javascript customization in ODK allowed an inclusion of longitudinal seed yam data entry, bi-directional synchronization, and on-device seed yam data management.

Parameters for seed yam tracking were also enforced in the XLSForm design. Forms were designed with the category of users in mind. The producers had forms that could be used to monitor certified seed; sources of the seeds, seed quantity, seed variety and seed costs. Crop management information (planting, weeding, fertilizer application, harvest) was also included in the forms.

To ensure continuity, apart from using the household ID as the primary key as with other software, the design included some data programming to ensure that once a bar code was read at planting, digits could be dynamically added to the bar code for each preceding stage by using the calculate option provided in the ODK XLSForms.

The pull data function was also used to transfer information on specific parameters of interest from one form to another to avoid a lot of time wasting in the field, which would require producers to feed in information each time there were new updates on crop management.

Some programming was also implemented to ensure that there was validation checks put into place. For example, the require field in the XLSForm was leveraged to ensure no skipping of questions, the calculate field was also implemented to ensure that data could be validated before proceeding to the next questions.

Seed regulators also had forms that were linked by bar codes and the ODK pull data function; one form was used to inspect farmers' fields while the second form was to get feedback from farmers on the inspection exercise. These forms had GPS locations which could be visualized on a map to show few details about the producer beforehand.

The results further indicated that respondents perceived the proposed system to be user friendly; this was their major expectation. The system could collect GPS data and handle multiple data types. It was found to be interoperable and cost effective.

5.2.2 Efficiency of the proposed prototype compared to other digital platforms and other data collection methods

As opposed to other digital platforms, ODK Aggregate, one of the suites of tools ODK offered efficient aggregation and storage for seed yam data. It provided visualization using maps and graphs and also allowed export and publishing of data in a variety of formats. The advantage of this tool was that the seed yam data could be hosted on cloud providers such as DigitalOcean, and Amazon Web Services, or an own local or cloud server, therefore providing data security and access of seed yam data in multiple locations.

The tool was effective in organizing seed production information, making it easier for institutions to monitor seed quality and certify producers, once data was collected it was pushed to the ODK aggregate which provided ready access to information, such as seed quantity, location, variety, geography, and availability. This information came in the form of raw data as well as visualization like maps and charts. The design also allowed integration to other software e.g Python and R which would provide users with more robust capabilities for seed yam data analysis and management.

With the use of the seed yam tracker, issues of nonresponse to sections of questions were reduced by the enforcement of validation of questions and inbuilt data quality checks that improve data completeness and accuracy. Users would therefore not be able to skip through questions as is the case with paper based surveys. Errors that may occur during data collation were also minimized since the transfer of data from the mobile devices to the ODK aggregate was seamless. The ODK Collect was linked to the Ona Aggregate server for easy uploads and retrieval to the data analysis software.

The seed yam tracker also potentially reduced the data collection and processing costs. These include many man-hours spent in locating seed producers, retrieving and verifying information from their previous stages of crop management, data entry and cleaning to improve the validity of research findings and also costs in printing of paper based questionnaires. (Maduka et al., 2017) compared the cost of ODK and paper based data collection. A cost analysis conducted showed the costs of deploying ODK on Android mobile phones for data management to be US $206.7 compared with the cost of paper-based data management of US $466.7 using already existing mobile phones.

As compared with other mobile data collection platforms, designing the prototype on the ODK platform further provided an immense and highly active community; a space where ideas on mobile data collection can be shared and discussed. It also offered both online and offline capabilities, which can sometimes be a hindrance for mobile data collection especially in Android based apps.

There was very limited literature on other seed tracking tools at the time of writing this thesis. The cassava seed tracker had been implemented by IITA and in use, having been built on the same principles as the seed yam tracker. There were also organizations such as the National Agricultural Seed Council (NASC) who were set to unleash their seed tracking app to tackle fake seeds and increase seed efficiency across the country. However, this app was also built on the principles of the yam seed tracker.

5.3 Challenges in implementing the Application

One of the challenges that may come with this application is data overload which could potentially crash the system. However, with advancement in technology users can store their data on cloud based services and native database devices to ensure information security. Another area of concern is user authenticity and accuracy of reported data, given that producers would feed in data without any supervision. This can however be mitigated by having all producers register when buying/receiving certified seed and thus allowing for follow up and verification of information. As the study is still ongoing, the initial results impress the value of using novel tools that promote developments in data collection and anticipate technological advances with significant gains in all areas of crop research and management.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study sought out to ascertain the challenges of the existing data collection systems in the yam sector in Ibadan. A number of challenges came up in this study including inaccessibility of seed yam data, lack of location information to help trace seed producers, lack of a streamlined system that could help in tracking certified seed based on all the stages of crop production, time delay in receiving feedback from seed producers, lack of standardized or validated data, translating into time consuming and costly data entry and cleaning, and the difficulty of accessing seed yam data that can help improve yam production.

System architecture was adopted from ODK and a custom design applied on the basis of the requirements and lessons learned through a study of the literature. The result was the seed yam tracker which was effective in resolving the challenges of the paper based methods that had been employed previously.

The seed yam tracker offered aggregation of data on the cloud which immediately provided access to data either in raw form or as summaries by use of charts and maps. The tracker also offered GPS information which assisted the seed regulators in locating the producers and in contact tracing. Through ODK's pull data function and multiple data types enabling, seed tracking through the various stages of crop management was made possible as well as referencing of previously fed in data. Through ODK's validation techniques, the tracker was able to offer validation checks to minimize errors in data and avoid non response.

Seed producers looking to increase their yield by applying improved technologies will be motivated to use the application even without incentives. The potential use of ODK in data collection in future will be pegged not only on the increased efficiency and accuracy, but also the reduction in costs from the use of affordable mobile devices.

6.2 Recommendations

The ODK system proposed for seed yam tracking is an important and effective solution that should be directed towards solving the identified constraints. Implementation of the system will streamline the process of seed yam data tracking all through the stages of crop management as well as inspection. The system will reduce greatly the time taken to provide information to stakeholders, it will act as an easier, user friendly and efficient system to provide feedback that ultimately informs decisions in the crop sector.

Scalability, an important aspect for continuity, was considered in the design and this meant that the prototype could easily be applied to other crops and not only specific to yam. The inbuilt algorithms of the seed yam tracker will also connect all key stakeholders, from seed producers, to regulators, traders, and extension services by providing a platform to view

Training on the use of the system may be important especially for seed producers in the rural areas, as well as provision in some cases of mobile devices that are compatible with ODK e.g. Android based smart phones/tablets. In order for any system to work in the agricultural sector, it is important as mentioned by some of the producers, to involve the government agencies in the process.

Mapping of seed producers will also become easier as well as crop disease surveillance because the system can collect a variety of question types such as GPS, bar codes, videos or audios. It is important to keep improving the system to address emerging needs in the yam sector. As such, integration of features is also a welcome approach.

6.3 Suggestions for further work

Based on the literature study, the field trials and the knowledge and experience gained through the process of testing the seed yam tracker application, several suggestions arise both to improve the system and also general suggestions provided by the respondents.

User-friendliness emerged as a critical aspect to adoption of mobile data collection systems. Inclusion of feedback mechanisms in addition to collecting data also emerged as one of the major concerns, especially by the producers of seed yam. To this end, it may be important to integrate SMS or email services to the system for easy communication.

In order to allow for easier navigation of mobile data collection systems, in App tutorials could be introduced which can be enabled on a need basis. Research could also be directed on better ways to perform disease surveillance using mobile technology. Most importantly, for any system to remain relevant, it should be visionary enough to anticipate the needs of crop research and management and incorporate solutions thereof.

REFERENCES

Aanensen DM, Huntley DM, Feil EJ, al-Own F & Spratt BG. (2009). EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *Plos one*, 4(9), 83-87.

Aidoo, Robert & Nimoh, Fred & Bakang, John-Eudes Andivi & Ohene-Yankyera, Kwasi & Fialor, Simon & Abaidoo, Robert. (2011). Economics of small-scale seed yam production in Ghana: implications for commercialization. *Journal of Sustainable Development in Africa*, 13(7), 65-76.

Aighewi, B. A., Maroya, N. G., & Asiedu, R. (2014). *Seed yam production from minisetts: A training manual*. Ibadan: IITA.

Anokwa, Y., Hartung, C., Brunette, W., Borriello, G. and Lerer, A. (2009). Open Source Data Collection in the Developing World. *Computer*, 42(10), 97-99.

Asumugha, G.N. and M.C. Ogbonna. (2013). *Promotion of improved yam minisett technology to improve productivity and reduce excessive use of food yam for planting in Ghana, Nigeria, Togo and Benin* (Report No. 2010/0022/SN/EA.PS/SE).CORAF/WECARD. Retrieved from http://www.coraf.org/database/projet/en/projet8.pdf

Baumüller, H. (2012). Technology Adoption among the Poor: The Role of Service Delivery through Mobile Phones. *SSRN Electronic Journal,* Working Paper Series 93, 9-11.

Brewer, Eric & Demmer, Michael & Densmore, Melissa & Honicky, R. & Pal, J. & Plauche, Madelaine & Surana, Sonesh. (2006). The Challenges of Technology Research for Developing Regions. *IEEE Pervasive Computing*, 5(2), 15 - 23.

Closs SJ. & Cheater FM. (1999), Evidence for nursing practice: a clarification of the issues *Journal of Advanced Nursing*, 30(1), 10-17.

Dillon, Brian. (2012). Using mobile phones to collect panel data in developing countries. *Journal of International Development* 24(4), 518 - 527.

Donner, J. (2008). Research approaches to mobile use in the developing world: A review of the literature. *The Information Society*, 24(3), 140–159

Fitzgerald, G. and FitzGibbon, M. (2014). A Comparative Analysis of Traditional and Digital Data Collection Methods in Social Research in LDCs - Case Studies Exploring Implications for Participation, Empowerment, and (mis)Understandings. *IFAC Proceedings Volumes*, 47(3), 11437-11438.

FAOSTAT (2014). *Crops data 2014.* Retrieved from http://www.fao.org/faostat/en/#data/QC

Frances Jeffrey Coker, Matt Basinger, (2010). *Open Data Kit: Implications for the Use of Smartphone Software Technology for Questionnaire Studies in International Development 2010.* Retrieved from htttp://qsel.columbia.edu/assets/uploads/blog/2013/06/Open-Data-Review-Article.pdf

Frontline (2021). *Reach everyone 2021*. Retrieved from https://www.frontlinesms.com/

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., and Borriello, G., (2010), Open Data Kit: tools to build information services for developing regions, Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (ICTD '10). *Association for Computing Machinery*, 18, 1–12.

Hoogeveen, J., Croke, K., Dabalen, A., Demombynes, G., & Giugale, M. (2014). Collecting high frequency panel data in Africa using mobile phone interviews. *Canadian Journal of Development Studies / Revue Canadienne D'études Du Développement*, 35(1), 186-207.

Hudson, Heather E. (2006). *From rural village to global village: Telecommunications for development in the information age*. LEA telecommunications series, Mahwah, NJ: Lawrence Erlbaum Associates.

Ironkwe, A. G., Asiedu, R., & Ekwe, K. (2017). Socio economic determinants of the adoption of yam minisett technology by women farmers in Abia state, Nigeria. *Journal of Sustainable Development*, 5(1), 82-90.

Javarosa (2020). *Javarosa wiki home 2020*. Retrieved June 01 2020, from https://bitbucket.org/javarosa/javarosa/wiki/Home

Jiggins, J. and Collins, K. (2003). Stakeholders and stake holding in social learning for integrated catchments management and sustainable use of water. *Thematic paper, Social Learning for the Integrated Management*.

Johnson, A. T. (2013). Exploring the use of mobile technology in qualitative inquiry in Africa. The Qualitative Report, 18(Art. 43), 1-14. Retrieved from http://www.nova.edu/ssss/QR/QR18/johnson43.pdf

Kenyon, L. (2009). Up-scaling sustainable clean seed yam production systems for small-scale growers in Nigeria. Final Technical Report. (Report No. R8416 (Za0648)). Retrieved from https://assets.publishing.service.gov.uk/media/57a08c4aed915d622c001251/R8416_FTR.pdf

Kwok, R. (2009). Personal technology: Phoning in data. *Nature*. 458 (7241), 959-61.

Maduka O, Akpan G, Maleghemi S. (2017). Using Android and open data kit technology in data management for research in resource-limited settings in the

Niger Delta Region of Nigeria: cross-sectional household survey. *JMIR Mhealth Uhealth*, 5(11), 171.

McNamara, K. (2003, December 9-10). Information and communication technologies, poverty and development: Learning from experience. infoDev Annual Symposium. Retrieved from http://documents1.worldbank.org/curated/en/741291468779079516/pdf/300760PAPER0ICT0Learning0from0Experience.pdf F

Mechael, P. (2009). The Case for mHealth in Developing Countries. *Innovations: Technology, Governance, Globalization*, 4(1), 103-118.

Mignouna, D. B., Abdoulaye, T., Alene, A., Asiedu, R., & Manyong, V. M. (2014). Characterization of yam-growing households in the project areas of Nigeria. *International Institute of Tropical Agriculture (IITA)*. Retrieved from https://www.researchgate.net/publication/283486398_Characterization_of_Yam-growing_Households_in_the_Project_Areas_of_Nigeria

Moser, C. O. (1998). The asset vulnerability framework: Reassessing urban poverty reduction strategies. *World Development*, 26(1), 1–19.

Muthiah, G., Prashant, S., & Jhunjhunwala, A. (2012). A review on challenges in implementing mobile phone based data collection in developing countries. *Journal of Health Informatics in Develop-ing Countries*, 6(1), 366–374.

Nokia (2021). *Creating technology that helps the world act together 2021*. Retrieved from https://www.nokia.com/

Open Data Kit (2021). *Open Data Kit 2021*. Retrieved from http://opendatakit.org/

openXdata (2007). *openXdata 2007*. Retrieved from https://arxiv.org/abs/2007.13889

Obayelu, A.E., Okoruwa, V., and Ajani, O. (2009). Cross sectional analysis of food demand in the North Central, Nigeria. *China Agricultural Economic Review,* 1(2), 173-193.

Rajput Z.A, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, Hartung C, Borriello G, Mamlin BW, Ndege SK, Were MC. (2011). Evaluation of an Android-based mHealth system for population surveillance in developing countries. *J Am Med Inform Assoc*, 19(2), 1-6.

RapidSMS (2016). RapidSMS. Retrieved November 11, 2018 from https://www.rapidsms.org/

Johnson, A. T. (2013). Exploring the use of mobile technology in qualitative inquiry in Africa. The Qualitative Report, 18(Art. 43), 1-14. Retrieved from http://www.nova.edu/ssss/QR/QR18/johnson43.pdf

Sahelcp.com (2014). *Nigeria yam improvement for processing*. Retrieved http://sahelcp.com/wp-content/uploads/2017/01/Nigeria-Yam-Improvement-for-Processing-ReportApril2014.pdf

Şahin, F., & Yan, Z. (2013). Mobile phones in data collection: A systematic review. International *Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, 3(3), 67-87.

Piette, S. J. John, Blaya Joaquin, Lange Lilta, (2011). Experiences in mHealth for Chronic Disease Management in 4 Countries, *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies* (pp. 1-5).

Udoh, Augustine & Idio, Anietie & Umoh, Emmanuel & Etuk, Unyime. (2008). Socioeconomic Factors Influencing Adoption of Yam Minisett Technology in South eastern Nigeria : A Probit Analysis. *Indian Res. J. Ext. Edu.* 8 (2&3), 1-4.

Verter, N., & Becvarova, V. (2015). An Analysis of Yam Production in Nigeria. *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*, 63(2), 659-665.

Vehovar, V., Berzelak, N., & Lozar Manfreda, K. (2010). Mobile phones in an environment of competing survey modes: Applying metric for evaluation of costs and errors. *Social Science Computer Review*, 28 (3), 303–318.

Waters, J. and Adger, W. N. (2017) Spatial, network and temporal dimensions of the determinants of adaptive capacity in poor urban areas. *Global Environmental Change*, 46, 42-49. Retrieved from https://ore.exeter.ac.uk/repository/bitstream/handle/10871/29622/WatersAdgerGEC2017.pdf;jsessionid=DF040ECBB9995B48A212219684634C8E?sequence=1

W3c (2021). *Leading the web to its full potential 2021*. Retrieved 01 March, 2021, from https://www.w3.org/

APPENDICES

Appendix I: Letter of introduction and consent

I am an MSc student currently undertaking Master of Science in Research Methods at Jomo Kenyatta University of Agriculture and Technology. I am conducting a research study to investigate the challenges of crop data collection and reporting systems. I also seek to evaluate the effectiveness of the prototype developed in regards to data collection, entry and recording of seed yam production fields. This will help improve aggregation of field data and real time visualization of data.

Your participation in this research study will go a long way in helping to improve the existing systems and will therefore be highly appreciated. Completion of the questionnaire should take no longer than 20 minutes. Your participation is confidential. All data obtained in this study will be reported as group data. No individual can be or will be identified. You may contact me with any questions by email (turry.ouma@gmail.com).

Appendix II: Questionnaire used for quantitative study part.

Section 1: General questions

1. What is your nationality?

   1. Nigerian 2. Kenyan

2 i) Have you ever worked in crop production?

1. Yes 2. No

ii) If your answer is Yes in (i) above, in which position? ....

1. Seed regulator/inspector
2. Producer
3. Researcher
4. Data collection/and use

3i) How is seed yam data predominantly collected in your country? ....

1. Paper methods 2. Using mobile devices 3. Both 4. I don't know

ii) How would you rate the efficiency of the method cited in 3 (i) above on a scale of 1-10?

iii) With the method in 3 (i) above, how long does it normally take to receive feedback?

1. Within a day

2. Within a month
3. Within 3 months
4. Within 6 months
5. 1 year
6. More than a year

4. To what extent is mobile technology in general, used in your country? ....

1. Wide 2. Intermediate 3. Low

5. Do you think mobile device applications could make seed yam data collection and reporting process more efficient? ....

1. Yes 2. No

6. Which kind of mobile phones are common in your country? ....

1. Smartphones. 2. Cell phones. 3. Both.

7. Which mobile platform do you use? ....

1. Android phones

2. Nokia phones

3. Blackberry phones

4. Windows phones

5. IPhone

f. Other phones

8. Do mobile networks cover the most part of your country? ....

1. Yes 2. No 3. I do not know

9. Is the crop information systems computerized in your country? ....

1. Yes 2. No

10. What are the main challenges that face crop information systems in your country?

A number of challenges were presented that face crop information systems in Nigeria.

11. How easy is it to access seed yam data on a scale of 1-10, 1 being very difficult and 10 being very easy?

12. What is the frequency of reporting seed yam data (monthly, weekly, quarterly, yearly)?

1. Within a day
2. Within a month
3. Within 1-3 months
4. Between 3-6 months
5. Between 6months-1 year

6. More than a year

13. How long does it take to collect and aggregate seed yam data using the current method (e.g., 1 day, 1 week, and 1 month)?

14. Who are the most common users of seed yam data in your country? (E.g. researchers, policy makers, seed regulators)?

Section 3: Prototype assessment

15. (Before users even look at the prototype) what do you expect to be able to do with the proposed prototype?

(Show the user the application and how to navigate through the pages. Allow them to freely explore the prototype)

16. After this exploration, do you think that this prototype would be useful to you? 1. Yes 2.No. Why or why not?

Prototype presentation assessment **Please rate the characteristics of the prototype on the left with the appropriate scale to the right:**

PROTOTYPE PRESENTATION CHARACTERISTICS	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
The prototype instructions are logical, clear, consistent and understandable for a layman	○	○	○	○	○
The prototype design, information content, and performance is up to par to collect and visualize results	○	○	○	○	○
The layout of text, graphics, figures, and links is user-friendly	○	○	○	○	○
The colors used in the prototype are appropriate	○	○	○	○	○
The text size and font are legible	○	○	○	○	○
The prototype provides an efficient and usable means of accessing, querying, analyzing, and interpreting data	○	○	○	○	○
The prototype addresses the	○	○	○	○	○

challenges of crop information systems					

17. The proposed prototype offers seed yam data mapping, charts and simple graphs for analysis. What are other features that could be included in future to improve data analysis and reporting process?

18. What in your opinion are the limitations of this prototype design with respect to data access, analysis, and interpretation?

19. How easy do you think it is to access data using this method of data collection: on a scale of 1-10?

20. How likely or unlikely would you use this product once it's finished?

21. Any other comments or suggestion.

Section 4: Improved aggregation and visualisation

Data visualisation assessment **Please rate the characteristics of the prototype on the left with the appropriate scale to the right:**					
DATA VISUALISATION CHARACTERISTICS	**Strongly Agree**	**Agree**	**Neither Agree Nor Disagree**	**Disagree**	**Strongly Disagree**

Data was presented in a logical and clear manner that was easy to interpret					
Data was visualised immediately after submission as long as there was internet connectivity					
Descriptive data was visualised in form of graphs, charts and tables					
The prototype would be useful in decision making for stakeholders					
The prototype facilitates the search, retrieval, sharing, analysis, and visualization of seed yam data					
The prototype addresses the challenges of crop information systems					

Appendix III: XLSForms for mobile data collection

I10 fx ${Select_Organization_name_if_applicable} = 'other__specify'

	A	B	
1	type	name	label
2	begin_group	grp_1	Seed Yam Production planning Form
3	note	image1	
4	note	note_begin	You are now beginning to fill the Yam Seed Production Planning form.
5	end_group		
6	begin_group	grp_1_header	Section I. Seed Producer Information
7	text	prod_code	Producer Code
8	select_one sq6eg48	Select_Organization_name_if_applicable	Select Organization
9	text	Name_of_Private_Company	Name of Private Company
10	text	Other_organization_type	Other organization type
11	text	name	Name of Seed Producer
12	text	sec1_mobile	Mobile number
13	text	sec1_email	Email
14	text	location	Location (Town / Village)
15	text	state	State
16	text	country	Country
17	geopoint	sec1_coordinates	Get the coordinates for the latitudes and longitudes
18	end_group		
19	begin_group	grp_seeddetails	Details for seed production and planting
20	select_one seed_purpose	sec1_seedpurpose	Class of Seed
21	text	otherpurpose	specify other
22	date	sec1_dateplanting	Expected date of planting/Propagation

Appendix IV: XLSForms for respondent mobile data collection

	A	B	C	D	E	F
1	type	name	label	hint	constraint	constraint_message
14	select_one nationality	nationality	1. What is your nationality?			
15	select_one answer	work_crop	2 i) Have you ever worked in crop production?			
16	text	crop_position	ii) If your answer is Yes in (i) above, in which position?			
17	select_one datacollected	data_collected	3 i) How is seed yam data collected in your country?			
18	integer	efficiency	ii) How would you rate the efficiency of the method cited in 3 (i) above on a scale of 1-10?			
19	select_one feedback	feedback_received	iii) With the method in 3 (i) above, how long does it normally take to receive feedback?			
20	select_one extent	mobile_extent	4. To what extent is mobile technology in general, used in your country?			
21	select_one answer	mobile_collect_efficient	5. Do you think mobile device applications could make seed yam data collection and reporting process more efficient?			
22	select_one mobphone	common_mobile	6. Which kind of mobile phones are common in your country?			
23	select_one mobplatform	mobile_platform	7. Which mobile platform do you use?			
24	select_one answer	network_coverage	8. Do mobile networks cover the most part of your country?			
25	end group					
26	begin group	section_2	Section 2: Evaluation of the proposed prototype			
27	select_one answer	computerized_system	9. Is the crop information systems computerized in your country?			
28	text	challenges_info_systems	10. What are the main challenges that face crop information systems in your country?			
29	integer	access-seedyamdata	11. How easy is it to access seed yam data on a scale of 1-10, 1 being very difficult and 10 being very easy			
30	select_one reporting	frequency_reporting	12. What is the frequency of reporting seed yam data (e.g. monthly, weekly, quarterly, yearly)?			
31	select_multiple data_users	users_seedyamdata	14. Who are the most common users of seed yam data			
32	end group					
33	begin group	effectiveness	Section 3: Effectiveness of prototype			
34	select_one answer	mobile_platform	15 a) We have designed a mobile platform to collect and visualize seed yam data; do you think this could improve the current way of collecting this data?			

survey | choices | settings