

https://dspace.org/

An offline mobile data capture module for health and demographic surveillance system (HDSS) studies

Baguiya, Adama

2016-11-04
http://hdl.handle.net/10539/21399

Downloaded from a DSpace repository. https://dspace.org/

AN OFFLINE MOBILE DATA CAPTURE MODULE FOR HEALTH AND DEMOGRAPHIC SURVEILLANCE SYSTEM (HDSS) STUDIES

Adama Baguiya

Student number: 899583



A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand in partial fulfilment of the requirements for the degree of Masters of Science

In Epidemiology (Research database management)

Supervisor: Gideon Nimako

Co-Supervisors: Dr Donatien Béguy

Bonface Nganyi

Johannesburg, May 2016

DECLARATION

I, Adama Baguiya, declare that the work contained herein is my own original work, and that where I have made use of others' ideas, I have referenced accordingly. This work has never been submitted before for any degree or qualification, certificate or publication.



Adama Baguiya

February, 25th 2016

University of the Witwatersrand, Johannesburg, South Africa

ABSTRACT

Considering the amount of data generated by the International Network for Demographic Evaluation of Populations and Their Health (INDEPTH), affiliated Health and Demographic Surveillance Systems (HDSS), and the complexity of their dynamic cohorts, introducing Information and Communication Technologies (ICT) and mobile technologies for data collection may significantly improve data quality and ease data processing and data management. So far, the level of use of mobile devices in public health research data collection and its determinants are not clearly known. We assessed the level of use of mobile devices for research data collection.

We conducted a cross-sectional survey in 29 HDSS sites out of 51, in Africa and Asia, to assess the current use of electronic/mobile devices for their core follow-up as well as the embedded studies. This survey revealed that the use of mobile devices was very (8 sites out of 29) low in the core HDSS follow-up. Meanwhile, a third of sites (34%) used mobile data collection for embedded studies. Motivations for using mobile data collection were data quality improvement and timeliness. On the other hand, devices' initial cost and unreliable internet connectivity were the major barriers to the use of mobile data collection. For those using paper-based data collection methods, Microsoft Access and Epidata were the two leading platforms for data entry whereas Research Electronic Data Capture (REDcap) was the most used electronic data capture system.

INDEPTH network is currently supporting the implementation of Open Health and Demographic System (OpenHDS) in HDSSs. This support can improve the rate of utilisation of mobile devices in HDSS data collection.

Electronic Data Capture (EDC) systems can potentially improve and facilitate data management in epidemiological studies. However, some of these systems lack mobile application or module

needed for field-based research. The second component of our work was to implement an open source mobile application, CTLS (Clinical Trials and Longitudinal Studies), that can be used with REDCap for electronic data capture. The application should be able to collect data offline and asynchronously upload onto REDCap database server or repository. We use REDCap in accordance with the findings of the survey: it was the most used EDC system. Such an offline mobile data capture feature will aid in overcoming the lack of mobile application and poor internet connectivity on the field, and improving the rate of utilisation of mobile data collection in remote settings. Our module has been tested using the entry and exit forms of the Nairobi Urban Health and Demographic Surveillance System in Kenya. It has been proved to be effective in fetching metadata from REDCap server, entering data offline and loading them to a remote database in an asynchronous fashion upon establishing internet connexion.

ACKNOWLEDGEMENTS

First, I would like to acknowledge, with many thanks, the scholarship from INDEPTH network for both the coursework and this research project for this Master's degree.

Secondly, I express my regards and gratitude to Prof Séni Kouanda, whose commitment to team building and strengthening of the leadership capacity of junior scholars has always helped me and other colleagues in building our careers. Many thanks for that.

I would also like to express my sincere thanks to my supervisors, Gideon Nimako, Dr Donatien Béguy and Bonface Nganyi. I owe them a deep sense of gratitude for their guidance and invaluable help during this research.

I am also grateful to the African Population and Health Research Center (APHRC), winner of the 2015 United Nations Population award, for hosting me during this research work. To all APHRC staff, thank you all for your support. I spent professionally fruitful time in the Statistics and Surveys Unit (SSU) and IT team.

I would also like to thank the ODK team for the great job they did before we could even think about this research. Their work has deeply inspired and supported ours.

To my wife Rébecca and son Ganill Williams, thank you all for giving me this strength and allowing me to go forward.

TABLE OF CONTENTS

LIST OF FIGURES VIII
LIST OF TABLES VIII
LIST OF ACRONYMS AND ABBREVIATIONS IX
CHAPTER 1: INTRODUCTION 1
1.1. BACKGROUND 1
1.2. PROBLEM STATEMENT 3
1.3. RESEARCH MOTIVATION 5
1.4. AIM AND OBJECTIVES 5
1.5. OUTLINE 6
CHAPTER 2: LITERATURE REVIEW 7
2.1. PARADIGM OF MOBILE DATA COLLECTION 7
2.2. MOBILE DATA COLLECTION PLATFORMS 8
2.3. RATE OF ADOPTION OF MOBILE DATA COLLECTION IN DEVELOPING COUNTRIES 9
2.4. ADVANTAGES AND OPPORTUNITIES IN MOBILE DATA COLLECTION 10
2.4.1. Ease of use 10
2.4.2. Cost-effectiveness 11
2.4.3. Timeliness 11
2.4.4. Data integrity 12
2.4.5. Data accessibility 12
2.5. MOBILE DATA COLLECTION CHALLENGES IN DEVELOPING COUNTRIES 12
2.5.1. Poor internet connectivity 12
2.5.2. Skills for setup, programming and maintenance 13
2.5.3. Mobile devices power supply 13
2.5.4. Data loss and system failure 13
2.5.5. Data security and ethical considerations 14
2.6. AVAILABLE OFFLINE MOBILE DATA COLLECTION APPLICATIONS 14
2.7. OVERVIEW OF CURRENT GAPS 15
CHAPTER 3: MOBILE DATA COLLECTION AT INDEPTH SITES 16
3.1. BACKGROUND 16
3.2. METHODS 17
3.2.1. Study setting 17

3.2.2. Study design, population and sampling 17
3.2.3. Data collection 18
3.2.4. Variables 19
3.2.5. Data analysis 20
3.2.6. Ethical considerations 21
3.3. RESULTS 21
3.3.1. Core HDSS data management 22
3.3.2. Characteristics of embedded studies 23
3.3.3. Mobile devices for data collection in embedded studies 24
3.3.4. Platforms used for data entry 26
3.3.5. Study characteristics and use of mobile data collection 27
3.4. DISCUSSION 28
3.4.1. Core HDSS data management 28
3.4.2. Use of mobile devices in embedded studies and motivations 28
3.4.3. Affiliation of principal investigator and mobile data collection 29
3.4.4. Data entry software and platforms 30
3.4.5. Limitations of the study 31
CHAPTER 4: OFFLINE MOBILE DATA CAPTURE MODULE RESEARCH METHODOLOGY 32
4.1. PROJECT DESIGN 32
4.2. SCOPE OF THE RESEARCH PROJECT 36
4.3. OVERVIEW OF THE PROPOSED OFFLINE DATA CAPTURE MODULE 37
4.3.1. Form provider 39
4.3.2. Offline data collection and storage 42
4.3.3. Asynchronous data upload 43
4.4. CONTRAST WITH PREVIOUS WORK 45
4.5. ETHICS, DATA ACCESS AND SECURITY 46
4.6. CHALLENGES AND LIMITATIONS 46
CHAPTER 5: EXPERIMENTAL SETUP AND VALIDATION 48
5.1. EXPERIMENTAL SETUP 48
5.1.1. Development, tools and technologies 48
5.1.2. Testing and prototyping 49
5.1.2.1. Form design 49
5.1.2.2. Data synchronisation 52

5.2. DATA SOURCES AND DATA PROCESSING METHODS 53
CHAPTER 6: CONCLUSION 55
REFERENCES 56
APPENDIX ONE: SENATE PLAGIARISM POLICY 60
APPENDIX TWO: ETHICS CLEARANCE 61
APPENDIX THREE: QUESTIONNAIRE DOWNLOADED FROM REDCAP 62

LIST OF FIGURES

Figure 1: A sample of questions on mobile data collection in HDSSs 19
Figure 2: Location of survey participants 22
Figure 3: Percentage of studies by reasons why they did not use mobile data collection by reason 26
Figure 4: ODK Collect main menu 33
Figure 5: Flowchart of activities in CTLS 35
Figure 6: Architecture of CTLS 38
Figure 7: CTLS main menu and root folder 40
Figure 8: Overview of Nairobi exit and entry forms REDCap data dictionary 40
Figure 9: Overview of XForm from Nairobi Urban HDSS entry form 42
Figure 10: Overview of redcap_data table 45
Figure 11: Forms on REDCap project interface 50
Figure 12: Forms as designed in REDCap 51
Figure 13: Example of CTLS text and date field 52

LIST OF TABLES

Table 1: Some popular platforms in health research mobile data collection 9
Table 2: Location of sub-Saharan African participating sites 22
Table 3: Characteristics of studies by type 24
Table 4: Differences in the use of mobile data collection by study characteristics 27

LIST OF ACRONYMS AND ABBREVIATIONS

ACID	Atomicity Consistency Isolation Durability
ADT	Android Development Tool
APHRC	African Population and Health Research Center
API	Application Programming Interface
CSV	Comma Separated Value
CTLS	Clinical Trials and Longitudinal Studies
DOM	Document Object Model
EDC	Electronic Data Capture
eHealth	Electronic Health
EHR	Electronic Health Record
EMR	Electronic Medical Record
GCP	Good Clinical Practice
GPRS	General Packet Radio Service
HDSS	Health and Demographic Surveillance System
HIS	health Information Systems
HTTP	Hyper-Text Transfer Protocol
HTTPS	Hyper-Text Transfer Protocol Secure
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
INDEPTH	International Network for Demographic Evaluation of Populations and Their Health
IP	Internet Protocol
JDBC	Java Database Connectivity
JDK	Java Development Kit
JSON	JavaScript Object Notation
mHealth	Mobile Health
MInCa	Mobile Research, Mobile information capture
ODK	Open Data Kit
OpenHDS	Open Health and Demographic System
OpenMRS	Open medical record system
PHP	Hypertext Pre-Processor
REDCap	Research Electronic Data Capture
SDK	Software Development Kit
SMS	Short Messaging System
SQL	Standardised Query Language
SSL	Secure Socket Layer
WHO	World Health Organisation
Wi-Fi	Wireless Fidelity
XML	Extensible Mark-up Language

CHAPTER 1: INTRODUCTION

In this chapter, we review the background of mobile data capture and explain the research problem to be addressed. Our motivation for this research and objectives are provided in the last sections of the chapter.

1.1. BACKGROUND

In public health, getting data of good quality on time has always been a concern. The quality of research data determines readiness for decision-making and capacity of health systems to respond efficiently in case of any population's health concern [1–3]. Disease surveillance, prevention and treatments are guided by information gathered from health information systems (HIS) and research studies, such as cross-sectional and longitudinal studies [2]. Policy-makers, researchers, governments' officials and programme managers all rely on data for defining populations' needs, prioritising them, setting goals, planning interventions, and monitoring progress. The real challenge is how fast research data can be accessed when real-time responsiveness to disease outbreaks (as well as other public health incidents) and continuous availability is required. For real-time research data management, research institutions need to rethink their research data infrastructure and processes.

Public health data can be drawn from two main sources. Firstly, it can be obtained from population-based investigations, such as censuses, surveys and cohort studies. Secondly, as an administrative entity, each and every component of a health system is also a source of administrative data. For instance, at the point of care, health workers routinely collect patients' medical data [4]. But, irrespective of the source, data always have to undergo data collection process [1,2]. In their conceptual framework of data quality assessment in public health practice, Chen et al. [1], highlighted the crucial role of data collection process as a key component of data quality assurance, to which further attention should be paid. Other dimensions of data quality such as completeness, accuracy, reliability, etc., are also strongly

related, and determined by the quality of data collection process. Weak data collection tools or methods negatively impact findings [5]. Hence, public health scientists and software developers continuously work on how to improve data collection tools, and propose better solutions to assist research data collection.

For many years, researchers and public health scientists relied on manual (i.e. paper based) data collection and processing methods [6]. However, over the past two decades, the advent of computers, and Information and Communication Technologies (ICT), have significantly changed the practice. Data can now be collected and captured using electronic data capture (EDC) systems. These systems allow users to collect data on papers and enter them on a computerised interface or form. For some studies conducted in health facilities, such as patients' records based studies, data can even be directly captured from the source into the system, eliminating data entry step. Recently, the penetration of mobile phones, smartphones, and other hand-held devices, have revolutionised health-related technologies for mobile healthcare, mobile data collection (mobile devices for data collection) and information dissemination [7]. Paper-based data collection is gradually phasing out [8], as the use of EDC systems significantly improves data quality [9,10].

In resource-constrained countries, mobile data collection tools are yet to be fully adopted. There are challenges to the adoption of both electronic data capture and mobile data collection [11,12]. Among these are skills for setup and maintenance, poor internet connectivity, cost of the hardware and software and acceptability are the most common challenges. Some of these challenges have been addressed over the past decade. There are numerous user friendly open-source software which help in reducing the cost of software and facilitating implementation. Some of the well-known open-source EDC systems are OpenClinica, OpenMRS, etc. REDCap is a freeware that offers advanced functionalities. It is however not regarded as an open-source

software because an agreement has to be signed with Vanderbilt University before its setup and use.

Regarding the lack of internet or poor connectivity, some developers proposed open-source Android-based mobile applications that allow offline mobile data collection and remote synchronisation with a database upon establishing internet connectivity [13,14].

1.2. PROBLEM STATEMENT

The number of research studies, including longitudinal studies and clinical trials carried out in low-income countries is increasing, and both sponsors and investigators are more and more demanding and are very interested in getting high quality data in a timely manner, with cost-effective solutions. The advent electronic data capture has revealed that paper-based data collection can actually be costly and time consuming [15]. Thus, whenever it is possible, researchers in developed countries are switching to mobile data collection. In addition, in clinical trials for instance, there are many data management requirements and standards that ensure compliance with Good Clinical Practices (GCP) and international guidelines [16,17]. EDC systems like OpenMRS[1], OpenEMR[2], REDCap[3], OpenClinica[4], Oracle clinical[5], etc., have features that can facilitate data management while ensuring compliance with those requirements. This gives a great opportunity of meeting the demand of high quality in the field of health research data management. However, in developing countries, the cost of software, hardware, human resources, etc., and many requirements of EDC systems can be challenging or even unaffordable. For instance, REDCap offers features such as randomisation, audit trail, de-identification and visit scheduling. This EDC system is widely used for longitudinal studies, cross-sectional surveys and clinical trials. Like the others mentioned above, Redcap is web-

[1] http://openmrs.org/
[2] http://www.open-emr.org/
[3] http://project-redcap.org/
[4] https://openclinica.com/
[5] http://www.oracle.com/us/products/applications/health-sciences/e-clinical/clinical/index.html

based. In addition, they do not all have mobile applications and cannot work with a different platform. Thus, there is a need for open-source mobile application programmed to link to the server side data store of these EDC systems. Further, as internet access in resource-constrained countries is challenging [16], mobile data collection faces both internet access and cost challenges, which hinder its adoption in clinical trials and longitudinal studies that require remote data collection.

These challenges can be overcome by using open-source application with offline mobile data collection capability linked to the server side data store of EDC systems. The Android development platform allows developers to design mobile applications that can download empty forms from a remote server via internet and store completed forms offline in a mobile devices memory. The forms can be sent to the server upon establishing internet connexion. ODK-Collect, the mobile application of Open data Kit is one of such. However, as opposed to the EDC systems mentioned above, ODK Aggregate, the server side of ODK, has been built for cross-sectional surveys, and does not offer all the features for health research studies, particularly longitudinal studies and clinical trials [18].

Our research which aimed to implement a mobile application called Clinical Trials and Longitudinal Studies (CTLS), based on ODK Collect, and link it with a selected EDC system can contribute in filling the gap and overcoming the challenge of limited resources (such as unreliable internet connectivity) in resource-constrained communities. In addition to that, the current lack of evidence on the rate of adoption of mobile devices for health research data collection is also a concern. Addressing it could help plan the promotion of the use of ICT and advocate for the use of mobile technologies in public health research.

1.3. RESEARCH MOTIVATION

Our research has been motivated, on one hand, by the lack of clear evidence regarding the rate and motivations of the use of mobile devices for data collection in health research institutions in developing countries. Some studies have explored this in sub-Saharan Africa, but they focused on electronic health records in healthcare facilities [18]. The World Wide Web Foundation also published a report in 2012 on mobile data collection in Africa. But, that report did not address the estimation of the rate of adoption. It rather gave important insights on key actors of mobile data collection, available technologies, business models and the unmet needs [19].

There is a need for open source mobile application that will be suitable for offline mobile data collection in longitudinal studies and clinical trials, and this project involves the implementation of an open source mobile application that works with selected open source EDC system commonly used in the field of health research. The application will allow offline data collection and ensure asynchronous storage to a remote electronic data capture back-end data store. We identified ODK-Collect [13,20], as a promising mobile application that can address this challenge. ODK-Collect is open source and implemented in a modular approach. As such, we will develop our module based on its source code to allow our application to communicate with REDCap database. Thus, researchers can use it for mobile data collection while all the rest of their study is managed by the REDCap desktop utility functions.

1.4. AIM AND OBJECTIVES

The aim of this research was to improve the use of mobile data collection in studies conducted in Health and Demographic Surveillance Systems, by proposing an open-source offline mobile data capture application.

The specific objectives of this research were as follows:

1. To assess the current use of mobile devices for data collection in Health and Demographic Surveillance Systems studies in Africa, Asia and Oceania.
2. To explore the various EDC system currently used in Health and Demographic Surveillance Systems within the INDEPTH network.
3. To develop and test an offline mobile data capture application for web-based EDC system based on ODK-Collect source code.
4. To analyse offline data ingestion into the database.

1.5. OUTLINE

This dissertation has been organised into six chapters. After this introduction (Chapter 1), Chapter 2 will present the literature review. The 3rd chapter describes the methods and presents the results of our cross-sectional survey on the rate of mobile data collection adoption in INDEPTH sites. In Chapter 4, we will describe our offline mobile data capture module, and Chapter 5 will describe its experimental setup, the test of the module as well as data access and security considerations and finally, the conclusion will be given in Chapter 6.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we review the current state of knowledge in the field of mobile data collection. This includes an overview of available tools and technologies.

2.1. PARADIGM OF MOBILE DATA COLLECTION

On a conceptual level, the application of information technologies in health care has introduced concepts such as telemedicine, Electronic Health (eHealth), Mobile Health (mHealth), Electronic Health Record (EHR), and mobile data collection. The definitions of some of these concepts are still not clearly standardised [21]. According to the World Health Organisation (WHO), eHealth is simply "*the use of information technologies for health*" [22]. Likewise, the use of mobile devices (PDAs, tablets, cell-phones and smart-phones) to support medical and public health practice is termed as mHealth [21]. With these definitions, Electronic health records (EHR) systems, also known as Electronic Medical Records (EMR), as defined by the Health Information Management Systems Society (HIMSS) [23], fits in the broader concept of eHealth. Electronic health records systems are normally setup within healthcare settings, such as hospitals and health centres, for managing patients' records. It usually stores patients' demographic data and information related to diagnosis, treatment and prognosis, using EDC systems such as OpenMRS, OpenClinica, and many other customised solutions developed by hospitals or countries themselves [24]. This does not necessarily imply the use of mobile devices.

Epidemiological research, public health and disease surveillance data collection using mobile devices, (mainly PDAs, mobile phones, tablets, laptops, etc.), fits in mobile data collection, and is what we will refer to as such. While EDC systems are usually designed as web applications linked to backend data store, mobile data collection uses a mobile application running on a tablet, a phone or PDA, linked to a cloud-based (hosted through the internet by Google, Amazon

or Microsoft, etc.) or an ordinary remote data store (i.e. a database in-house), allowing data to be collated remotely on a real-time basis from the field [13].

The functionalities of mobile devices have improved in the past 2 decades. Twenty years ago, most of mobile data collection or other components of mHealth, were using simple mobile phones and GPRS connectivity [25]. Data were sent in textual format using Short Messaging System (SMS) [14]. Nowadays, smartphones and tablets are increasingly taking over, and they offer sophisticated features. Most of the current devices use third or fourth generation cellular connectivity and Wi-Fi. Further, they can handle Global Positioning System (GPS) data, images, barcodes, fingerprints, etc. This brings more opportunities to the landscape of health data collection.

2.2. MOBILE DATA COLLECTION PLATFORMS

Mobile data collection is an improved method for data capture and storage. They involve devices such as PDAs, cellular phones, smartphones, tablets, netbook and notebook [14]. Currently, on the market, there are many types of hand-held devices running different operating systems. The most used and well known are Android, iOS for iPhone and iPad, Windows mobile, and Blackberry [26]. In Table 1, we describe the most popular platforms for mobile data collection. Most of them use WiFi connectivity and they all use 3G. Further, both proprietary and open source platforms can handle all data types (Table 1).

Table 1: Some popular platforms in health research mobile data collection

Tool	License type	Mobile compatibility	Data type	Network protocol	Storage
Atlas.ti mobile	Proprietary	Android tablets and iPad	Text, images, audio and video	GPRS, 3G, WiFi	Hosted
CSEntry CS pro data entry	Open source	Android phone and tablet	Text	GPRS, 3G, WiFi, bluetooth	Local
EMIT	Proprietary	Java phones	Text	GPRS	Hosted
EpiCollect	Open source	Android and IOS phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	hosted
FrontlineSMS	Open source	Basic phones	Text (SMS)	GSM	Local
GIS mobile data collection	Open source	Android and IOS phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Hosted
KoboCollect	Open source	Android and IOS phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Hosted or local
Magpi	Proprietary	Android and IOS phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Hosted
Mobenzi researcher	Proprietary	Java, BlackBerry and Android phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Hosted
ODK Collect	Open source	Android smartphones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Local or hosted
Pendragon	Proprietary	IOS and android and Windows phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Local and hosted
RapidSMS	Open source	Basic phones	Text (SMS)	GSM	Local
SurveyCTO	Proprietary	Android phones and tablets	Text, Video, Audio, images, GPS, Barcodes	GPRS, 3G, WiFi	Hosted

Adapted from Shao D, 2012 [14]

2.3. RATE OF ADOPTION OF MOBILE DATA COLLECTION IN DEVELOPING COUNTRIES

The world overall rate of penetration of mobile devices have significantly increased, to the extent that, there are nearly as much cell phones as individuals in the world. Meanwhile, the gap between developed and developing countries remains wide [27]. In 2013, the rate of penetration of mobile phones was 128% in developed countries, 89% in developing countries, and 63% in Africa [28]. This high cell phones penetration rate is considered as a factor that arouses the use of hand-held devices for data gathering in Africa [13]. However, the use of mobile devices for health research or public health data collection seems low in developing countries, and studies on this topic are scarce [11]. According to Tomlinson et al, in developing

countries, “the use of mobile technology as a research instrument is in its infancy” [29]. Aranda-Jan et al, in a systematic review, searched Pubmed and OvidSP for articles on mHealth published from 2003 to 2013. Only 10 papers were found on data collection/transfer and reporting, and 4 on disease surveillance and intervention monitoring [11]. Most of the studies published on the rate of adoption of ICT in health in developing countries targeted EHR (not mobile data collection). Akanbi et al. [18] for instance, explored the EHR in sub-Saharan Africa and reported on progress and challenges. Boyera et al, published in 2012, on behalf of the World Wide Web foundation, a report entitled “*Mobile data collection in Africa*” [19], which focused mainly on available technologies and guidance for their use, than the level of use. Yet technologies and tools are available and are extensively presented in the literature. However, this does not answer the question about the proportion of studies in developing countries that opt for mobile data collection in the context of high mobile devices penetration.

2.4. ADVANTAGES AND OPPORTUNITIES IN MOBILE DATA COLLECTION

Opting for mobile devices for health research data collection comes with many advantages and opportunities already identified by previous studies [9,12,29].

2.4.1. Ease of use

The use of mobile devices is less cumbersome and more user-friendly than manual data collection [29]. Although some researchers tend to train field workers on how to use mobile devices [30], data collectors usually get progressively and easily familiar with the device and the questionnaire, and no additional time is required. Knipe et al, reported that in their experience of using PDAs in a large cluster randomised control trial in Sri Lanka, data collectors that had no previous experience with PDAs, used just two days to get acquainted with the devices and the software, and the duration of the survey for each individual decreased from the second week [12]. Tomlinson et al, in South Africa, also worked with community health

workers who had no previous experience as data collectors, but they found the process suitable [29].

Regarding the questionnaire design, many tools facilitate the user's experience in implementing mobile data collection without any software development background. For instance, both Magpi and ODK have user interfaces for designing forms with clicks or drag-and-drop functionalities [13,31].

2.4.2. Cost-effectiveness

We often approach cost-effectiveness of mobile data collection by comparing it to paper-based data collection. At least, using mobile devices for data collection is not significantly more expensive than using a paper-based system for a small scale study as shown by Giduthuri et al. [30]. The initial cost of devices is what usually makes electronic device solutions slightly more expensive than manual data collection. However, it is proved that for a large scale studies, tablet-based data collection definitely becomes cheaper than the traditional method [30]; Thriemer et al., reported 25% cost reduction with PDAs in Zanzibar [10]. Moreover, irrespective of the duration of the study, with mobile data collection, the total cost of ownership is lower as mobile devices and software can be reused for a long period of time over several studies.

2.4.3. Timeliness

Data is entered straight from the field by data collectors, and can be checked from a server on a daily basis. Data is thus almost immediately available from data stores because it merges data collection and data entry steps of the research data lifecycle [32]. This is one of the major advantages of mobile data collection.

2.4.4. Data integrity

In general, mobile data collection uses smart forms (XForms). Those forms allow check controls and skip logics at the application level. In addition, as data hit the database in real-time, some other constraints can be programmed to facilitate supervision and monitoring of inconsistencies and data quality [29]. This can be facilitated by a dashboard or an interface often proposed along with the application.

2.4.5. Data accessibility

Many mobile data collection solutions or software provide an opportunity for managing access to the data store by allocating roles and privileges. These features can improve control over the use or manipulation of the data. For instance, a sponsor or partner can be allocated a view privilege, and therefore will remotely follow the progress of the work but will not be able to edit stored datasets. Meanwhile, data collector will be allowed to enter or edit data but not view previously collected data. Mobile data collection offers an opportunity for accessing data almost in real-time as it is uploaded from the field to the server [13].

2.5. MOBILE DATA COLLECTION CHALLENGES IN DEVELOPING COUNTRIES

The use of mobile technologies in health sector has its global challenges that institutions encounter regardless of whether they are in a developed or a developing country. Firstly, it comes with changes to be applied on the current system. Therefore, it needs some goodwill to embark on those changes [33]. In addition, some other aspects can reveal technical challenges.

2.5.1. Poor internet connectivity

Most of the applications use internet connectivity (cellular data or Wi-Fi) to send data to their data store. Many studies, particularly in rural settings, have mentioned poor internet connectivity as a challenge in mobile data collection [34].

When data managers have to opt for a specific data collection systems, they always have to take into consideration all available options and how they operate. In settings with poor internet connectivity, web-based solutions are very difficult to implement, although they provide a lot of advantages. Consequently, they are rarely explored because they raise a serious concern about how to link with a remote database. This adds to hacking and malicious users concerns that always come along with the use of internet. These can be mitigated with proper access right and privileges management, use of firewall, encryption, activity logging and regular monitoring of log files, etc.

2.5.2. Skills for setup, programming and maintenance

In some countries, the implementation of mobile data collection have often involved programming or other technical skills from overseas [12]. This has often been reported as a factor that makes researchers face delays, spend more time and money for system setup and maintenance. For a long time, researchers have been under impression that sub-Saharan African countries lack skills for mobile data collection. But evidence suggests that it is not really a concern. For instance, in Sri Lanka, Knipe et al reported that the devices used for their studies were programmed and maintained abroad. They realised afterwards that they could have gotten skills in Sri Lanka for that job and prevent delays [12].

2.5.3. Mobile devices power supply

Mobile devices batteries are known to drain out quickly [35]. Some batteries last less than 10 hours [12]. But carrying a spare battery and making sure it is charged as permanently as possible can easily overcome this challenge. Field workers can even carry dual chargers [12,34].

2.5.4. Data loss and system failure

The device can get misplaced, damaged, stolen or lost [34]. Moreover, hardware or software failure can also occur and cause data loss. Luckily, such cases do not occur frequently.

Tomlinson et al, in South Africa, in 2009, did not experience any system failure or data loss throughout their study that involved 39 665 households [29].

2.5.5. Data security and ethical considerations

Electronic data handling raises concerns regarding data security and individuals' personal information protection. Many researchers used encryption to securely send data from a node to another [29,36]. Passwords, antiviruses and firewalls can also help reach a comfortable level of data security and protection [29].

2.6. AVAILABLE OFFLINE MOBILE DATA COLLECTION APPLICATIONS

There have been many attempts to address the challenges raised in the use of EDC systems. Many of such systems are suitable for epidemiological studies but require permanent internet connectivity. Previous works have explored how to leverage it to fit their needs or simply improve certain functionalities. The ODK team proposed a feature of ODK-Aggregate called ODK-aggregate publisher, which can be used to publish ODK-Aggregate data into REDCap. However, data flow with this publisher goes first to ODK-Aggregate before publishing or streaming data into REDCap server. Unfortunately, this is still in alpha and it does not preserve groups and repetition [37].

Besides ODK team, some other research used ODK to develop applications as it is open source. OpenMRS has been linked for interfacing with an Android based mobile application called AccessMRS by Fazen et al. These authors also used ODK source code and designed smart forms to interface OpenMRS for the use of community health volunteers in Kenya [36]. Other applications such as EpiCollect+, Mobile Research, Mobile information capture (MInCa), Afla Cohort study, KoboCollect, etc, provide offline mobile data collection capacity independently (Mobile research, MInCa) or based on ODK either with their own backend data store (KoboToolbox, EpiCollect, etc.…) or with ODK Aggregate (Afla Cohort study). In addition to

these applications, MDbot proposed a REDCap mobile application that runs on iPhones and iPads [38]. More recently, REDCap announced their own mobile application [39]. But, REDCap is a freeware and is different from our module which is based on the principle of link between 2 platforms: an open source mobile application and REDCap server.

2.7. OVERVIEW OF CURRENT GAPS

From the review above, we can summarise the gaps regarding mobile data collection tools availability as follows:

- Gap 1: lack of evidence regarding the level of adoption of mobile devices in health research data collection in developing countries in general. There is no available study or review on the level of use.
- Gap 2: Lack of open source mobile application for epidemiological and public health research web-based EDC systems. No application currently gives the opportunity of working across platforms (EDC system and mobile application) for health research data collection.
- Gap 3: Lack of integration between epidemiological and public health research web-based data capture systems and available open source mobile application that have offline storage and asynchronous data synchronisation capability.

CHAPTER 3: MOBILE DATA COLLECTION AT INDEPTH SITES

In this chapter, we report the results of our cross-sectional survey, conducted in INDEPTH affiliated HDSSs. After a background section, we describe our methods and report the results. Main findings will be discussed in the last section of the chapter.

3.1. BACKGROUND

Hand-held devices have increasingly penetrated low and middle income countries over the last decade [25,40], and this will continue over the next 5 years [41]. Although some studies have reported controversial debate regarding the added value of the use of mobile technologies in health [12,42], two things are now considered as facts. First, mobile devices have significantly contributed in health related activities such as healthcare, public health intervention and surveillance, community health workers' activities, health research, etc. Secondly, mobile data collection is faster, easier and more accurate. Other studies also proved that it is furthermore cheaper [10]. Similar to other fields, telephone and tablets use in health research data collection is now an option to be considered whenever one has to conduct a field-based research.

However, in resource-constrained settings, despite the fact that mobile data collection is more and more used, there is no evidence about the actual rate of penetration of hand-held computers and smartphones in research data collection. Some studies have mentioned that the use of mobile devices is low in developing countries and have explored the reasons that impede their utilisation [12]. Unfortunately, the level of use by researchers for data collection is still not quantified. In this chapter, we intend 1) to determine the proportion of studies for which researchers have opted for mobile devices for data collection in HDSSs, 2) to identify the main reasons why they do not use it for some studies, and their motivation when they do.

3.2. METHODS

3.2.1. Study setting

This assessment was conducted in HDSSs affiliated to INDEPTH network. Overall, 51 HDSSs were operational as at May 2015, in Africa, Asia and Oceania.

HDSSs are research platforms located in 20 low and middle income countries (Africa, Asia and Oceania) as defined by the World Bank [43]. Most of them are either entirely or partially located in rural areas. Only Nairobi Urban HDSS and Ouagadougou HDSS are entirely urban.

HDSSs follow up a dynamic cohort with regular visits (at least once or twice a year), to capture health and demographic events within a specific population under surveillance. The core events followed in the HDSSs are: pregnancies, births, morbidity, unions, migration, and deaths [44]. In addition to their core follow-up, HDSSs regularly carry-out embedded studies in their sites, and generate important amount of data. However, with regard to both the embedded studies and the core HDSS follow-up, data management methods are not standardised across sites. While some sites frequently use mobile devices, some find paper-based data collection methods more suitable for their research activities. This heterogeneity impedes the potential of data integration, data sharing across site and multisite studies.

3.2.2. Study design, population and sampling

We conducted an online survey from May 14th, to June 20th, 2015. The study was to examine the use of mobile devices for HDSSs core modules and embedded studies during the period of the survey. We listed and invited all 51 HDSSs from Africa, Asia and Oceania, to participate in this survey.

Regarding the embedded studies, we included in our sample the 5 most recent studies that were ongoing in the HDSSs. This was to prevent the bigger sites that have many embedded studies carried-out with either paper-based or mobile data collection, to influence the results. Even

when a site had many ongoing studies, we only included the five most recent considering their dates of start. This also helped increase the response rate by minimising the length of the questionnaire.

3.2.3. Data collection

During this survey, we used an online questionnaire designed in REDCap. The questionnaire was deployed online, from May 14th to June 20th, 2015. REDCap has a feature that allows users to invite participants to fill questionnaires using their email addresses. This sends an email with a link to the questionnaire to participants. We used that functionality to send our questionnaire to HDSSs data managers.

Prior to sending the questionnaire, we contacted the site leaders and requested permission to include their site, and the contact details of the data manager who can respond to the questionnaire. After receiving the consent from the site leader, the link to the online questionnaire was then sent to the data manager.

Upon completing the questionnaire, a button allowed respondents to submit their form by a click. On clicking that button, the responses get uploaded into our database. Reminders were automatically programmed to weekly remind data managers until the form is completed and submitted. Partially completed forms could be saved with a unique code for later access, and thus, the reminder will keep being sent to the respondent to notify him/her that the questionnaire is not complete. Questionnaires could be accessed and edited using the re-entry code, as each link was unique for the site to which it was sent (the email address of the data manager was used to identify the sites). After submission of the form, only the study coordinator could edit responses; respondents were not allowed access to it anymore.

To keep the form easy to fill and minimise the amount of text respondents had to type, we designed all the questions (except the option "other" that needed to be specified) to be either radio button, dropdown menu, or check-box as illustrated in Figure 1.

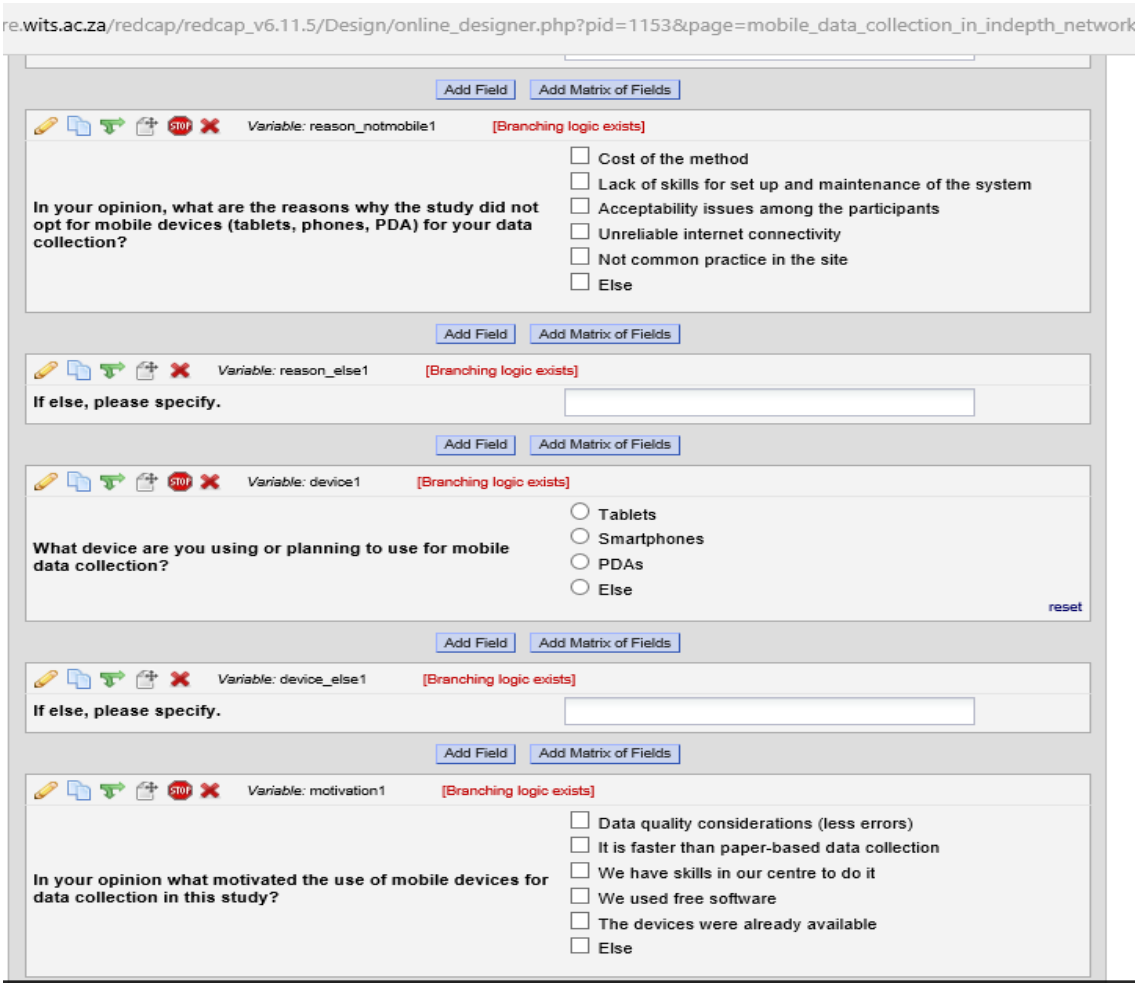

Figure 1: A sample of questions on mobile data collection in HDSSs

3.2.4. Variables

The variables of interest in this assessment were:

- Use of mobile devices for core HDSSs data management (yes/no) ;
- Use of mobile devices for embedded studies data management (yes/no);
- Type of EDC system used for the core HDSS data entry;
- For sites that use paper-based data collection methods, we sought information about which electronic data capture system they use for data entry;

- Number of embedded studies conducted in the HDSS. Although specific information about the studies were asked for the 5 most recent studies, we reported the total number of studies carried-out in the site during the study period;
- The source of the study funds. This was a binary variable for which we asked whether funds used for the study was from a grant won by the site or it was self-funded;
- The institution of origin of the principal investigator. We asked whether the principal investigator was from the site or not. If not, respondents were asked to specify his site or centre or university/institution of origin;
- Reasons why mobile devices were used. For studies that used mobile devices, we asked their motivation of this decision;
- Reasons why mobile devices are not used. For those who did not use, we also asked the reasons why they did not opt for mobile devices.

3.2.5. Data analysis

We performed data analysis using Stata 13, and the following descriptive and summary statistics were computed:

- Proportion of longitudinal studies in which mobile data collection has been used by study location (urban, rural or both), by origin of the fund and by origin of the principal investigator;
- Proportion of use of EDC systems by type of system;
- Proportion of utilisation per type of mobile device;
- Reasons why sites are not using mobile devices: percentage per type of reason
- Reasons why sites are using mobile devices: percentage per type of reason

We used Chi-square and Fisher's exact test for bivariate analysis. Chi-square test was for comparing proportions and the Fisher's exact test was applied when we encountered small expected values.

3.2.6. Ethical considerations

During this assessment, we paid attention to ethical considerations such as informed consent and authorisation from sites' leaders and, right of refusal to participate without any risk or harm, and protection of confidentiality of collected information. Thus, prior to deploying the questionnaire, we sent our information sheet and all other authorisation documents to, and sought consent from, each site leader. After obtaining the consent, we contacted the data manager and sought their consent as well.

No private or sensitive information was recorded from data managers. We did not store data managers' personal information either.

3.3. RESULTS

Overall, we invited 51 site leaders to participate in the survey. Thirty-nine of them (76.5%) responded and agreed to enrol their site. Among those 39 potential participating sites, 29 (69.2%) sites' data managers responded to the questionnaire.

Majority of the participating sites were from West-Africa (10 out 29) and East-Africa (8 out 29). Seven of them were from Asia. Only one site from Oceania participated in the survey (Figure 2).

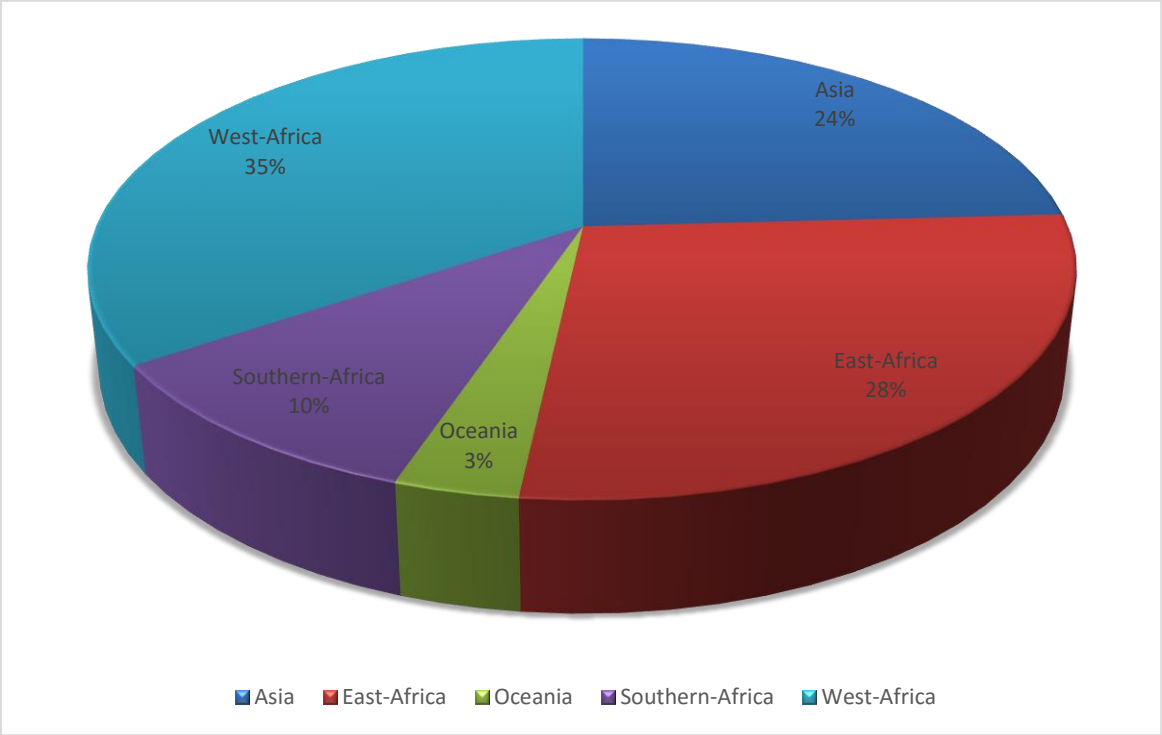

Figure 2: Location of survey participants

Twenty-one countries were located in sub-Saharan Africa. Burkina Faso and Kenya are the most represented countries with respectively 4(19%) and 3(14.1%) sites as shown in Table 2.

Table 2: Location of sub-Saharan African participating sites

Country	Number of sites	Percentage
Burkina Faso	4	19.0
Ethiopia	2	9.5
Ghana	1	4.8
Guinea-bissau	1	4.8
Kenya	3	14.1
Malawi	1	4.8
Nigeria	2	9.5
Senegal	2	9.5
South Africa	2	9.5
Tanzania	1	4.8
Uganda	2	9.5
Total	**21**	**100.0**

3.3.1. Core HDSS data management

Among the 29 sites, 8 were using mobile devices for data collection for their core HDSS follow-up. Majority (18 out of 29) of them were using paper-based data collection methods, and 3 sites reported the use of laptops or netbooks on the field. Those who used paper-based data collection

or laptops/netbooks mainly use their own platforms (13 out of 21), built by the programmers from their own site for data entry.

3.3.2. Characteristics of embedded studies

In total, 69 embedded studies have been included in this survey. Sixty percent (n= 39) of them used paper-based data collection methods and 34% (n=21) used mobile devices. With regard to the study settings, half (50.8%) of the studies was conducted in rural areas and 34% in both urban and rural areas. For 4 out of 5 studies (80%), the principal investigator was from the HDSS where the study was conducted.

Table 3 presents the percentage of various characteristics of the embedded studies, such as field of study, location, affiliation of the principal investigator and the funds, duration and data collection method by design.

Table 3: Characteristics of studies by type

	Type of study					
	Longitudinal		Cross-sectional		Total	
Variables	n	%	n	%	N	%
Field of study (n= 68)						
Health system research	8	15.4	5	31.2	13	19.1
Maternal and child health	8	15.4	4	25.0	12	17.7
HIV surveillance, prevention or treatment	3	5.8	0	0.0	3	4.4
Malaria surveillance, prevention or treatment	8	15.4	2	12.5	10	14.7
Tuberculosis and other diseases surveillance, prevention or treatment	11	21.1	4	25.0	15	22.1
Clinical trial	6	11.5	-	-	6	8.8
Other	8	15.4	1	6.3	9	13.2
Study location (n= 65)						
Rural	24	47.1	9	64.3	33	50.8
Urban	5	9.8	2	14.3	7	10.8
Both rural and urban	22	43.1	3	21.4	25	38.4
Duration of the study (n= 64)						
12 months or less	4	7.8	5	38.5	9	14.1
1 to 2 years	11	21.6	2	15.3	13	20.3
2 to 5 years	28	54.9	5	38.5	33	51.5
More than 5 years	8	15.7	1	7.7	9	14.1
Multisite study (yes)	24	46.2	5	31.3	29	42.7
Multi-country study (yes)	18	34.6	2	13.3	20	29.9
Principal investigator from the site (yes)	42	80.8	12	80.0	54	80.6
Origin of fund (n= 66)						
Self-funded study	8	15.4	3	21.4	11	16.7
Grant	44	84.6	11	78.6	55	83.3
Data collection methods (n= 65)						
Paper-based	29	55.8	10	76.9	39	60.0
Mobile data collection	21	40.4	1	7.7	22	33.8
Other methods	2	3.8	2	15.4	4	6.2

3.3.3. Mobile devices for data collection in embedded studies

- *Rate of use of mobile devices*

Among the 69 studies included in this survey, 22 (34%) used either smartphones, tablets or PDAs for data collection. Only one of these was cross-sectional survey. Tablets were the most used devices; 11 studies out of the 22 used tablets and 4 used smartphones.

- *Data transfer to the server*

Fifteen of the 22 studies used internet connexion to upload their data to their data store. Ten of them exclusively used internet and the other five studies used both a local network and the internet for data transfer.

- *Motivation for using mobile data collection tools*

The use of mobile data collection method was motivated by data quality improvement considerations for 16 of the studies, and timeliness for 19.

In 16 of the 22 studies, data managers opted for mobile devices because they considered that it will guarantee both data quality and timeliness.

For 11 of the studies, the motivation was the fact that they had skills to do so, and for some other 6 studies, they used it because the software for data collection was free.

- *Reasons for not using mobile devices*

Overall, 43 (66%) studies did not use mobile devices for data collection for various reasons. The cost of devices and software, lack of skills and unreliable internet connexion were the most reported reasons why researchers did not opt for mobile data collection. Six studies reported that the procedure of the clinical trial they were conducting prevented them from using mobile data collection. Auditors of the studies needed paper-based forms as data sources for audit purposes.

Percentages of the studies for which data managers reported reasons as per why they did not use mobile devices for data collection, are presented in Figure 3. Forty percent of the studies reported that they did not use mobile data collection because of the cost of the method. Lack of skills for setup and maintenance of the system and unreliable internet connectivity were the reasons why researchers did not opt for mobile data collection in respectively 35% and 30% of

the studies. This method was not a common practice some sites, and that was why it was not explored in 25% of the studies (Figure 3).

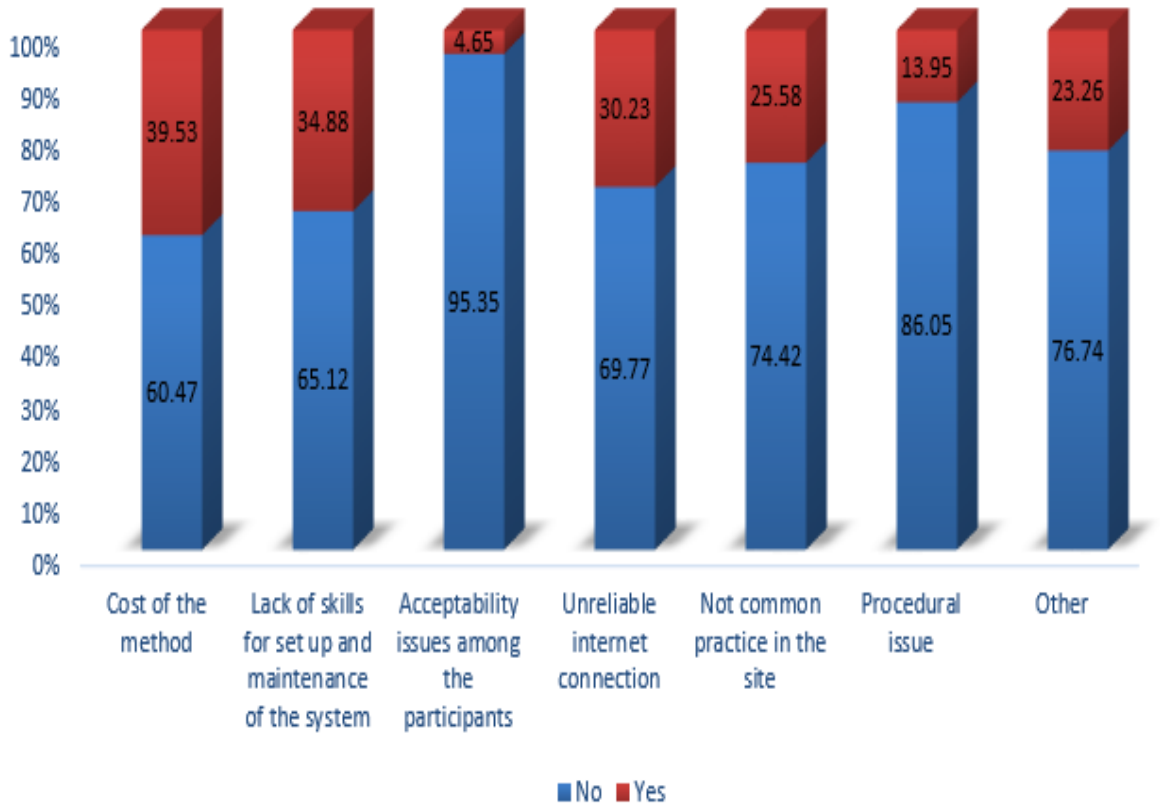

Figure 3: Percentage of studies by reasons why they did not use mobile data collection by reason

3.3.4. Platforms used for data entry

Data entry platforms were reported for the 29 studies that were conducted with manual data collection method. Twelve of them used sites own platforms (developed by developers from the site). Seven studies used Microsoft Access, four of them used Epidata, and 4 other used Cs-Pro. FoxPro has been used by 3 studies.

Among the available web-based electronic data capture system, only REDCap and OpenClinica have been used respectively by 3 and 2 studies.

3.3.5. Study characteristics and use of mobile data collection

When we compare the percentage of use of mobile devices in categories of different characteristics of the studies as factors in a bivariate analysis, there seemed to be some patterns. Although the differences were not strongly significant, the type of study (p =0.046), affiliation of the principal investigator (p=0.087) and the study location (p=0.088) seemed to have the potential of being associated with the implementation of mobile data collection (Table 4). Stronger studies could seek enough power to better explore this with a multivariate approach.

Table 4: Differences in the use of mobile data collection by study characteristics

	Use of mobile devices (Yes)		
Variables	N	Percentage	p-value
Type of study			
Longitudinal	52	40.4	
Cross-sectional	13	7.7	0.046[b]
Use of mobile devices for core HDSS			
No	44	27.3	
Yes	21	47.6	0.160
Multisite study			
No	37	35.1	
Yes	28	32.1	0.801
PI[a] from the site			
No	12	58.3	
Yes	53	28.3	0.087[b]
Origin of funds			
Self-funded	11	54.5	
Grant	54	29.6	0.162[b]
Duration of the study			
Two or Less than 2 years	22	27.3	
More than 2 years	41	36.6	0.455
Location			
Rural	32	21.9	
Urban	7	57.1	
Both rural and urban	25	44.0	0.088[b]

a Principal investigator
b Fisher's exact test

3.4. DISCUSSION

3.4.1. Core HDSS data management

We found that most of the HDSSs (18/29) used paper-based data collection methods for their core follow-up. When it comes to data entry, they mostly used in-house platforms. These results suggest that although sites have skills to develop their own platforms, they are not yet fully embarked on mobile data collection for the core HDSSs follow-up. HDSSs follow complex health and demographic events over time, and deal with temporal relational data which is somehow hard to manage. Sites seemed ready to use mobile devices for smaller studies than to venture into mobile data collection for their dynamic cohort, in which data is repeatedly and regularly collected every 4, 6 or 12 months. Furthermore, many sites started their activities many years before the increasing penetration of mobile technologies and their application into research data collection. Thus, migrating to mobile devices comes with its burden in terms of data migration, changes in data management plans, infrastructure, equipment, skills, and time. Hence, one can be reluctant to move to mobile devices after considering all these aspects.

The fact that the sites use in-house platforms for data entry could mean that they have or can easily get the required human resources for handling mobile data collection technologies. Furthermore, some applications, already developed and suitable for HDSSs, are available. Open Health and Demographic System (OpenHDS)[6] for instance, is an open source Java based application, fully developed for HDSSs data model, and represents an opportunity that can be explored and adopted by sites.

3.4.2. Use of mobile devices in embedded studies and motivations

In our study, 22 studies (34%) was conducted using mobile devices for data collection. There is a lack of evidence on the rate of use of mobile devices for research data collection in low

[6] https://code.google.com/p/openhds/

income countries. Some studies explored the use of electronic health record systems but, there is limited evidence so far on the level of use of mobile devices in health research data collection.

The main reasons of failure to use mobile data collection were: the cost of the method and the fact that internet connexion is not reliable enough for data managers to feel comfortable in opting for mobile devices. Regarding the motivation, data quality and timeliness were the major reason why data managers opted for mobile devices. These motivations are currently well known as part of the advantages of mobile data collection [29,32]. Likewise, the reasons why, in our study, manual data collection were sometimes preferred over hand-held devices, were equipment and setup, which were also common to both electronic data capture systems and mobile data collection systems implementation in resources-constrained countries. One has to acknowledge that to change from manual data collection to mobile data capture, there is an investment to bear in terms of equipment and infrastructure setups. Open source solutions can significantly decrease the financial burden of such system [10].

Concerning internet connectivity, it is established that a data collection system that entirely relies on internet connectivity is not reliable in developing countries. However, previous studies have proved that, with mobile devices, one does not have to be continuously connected to the internet. Solutions exist to temporarily store data in phones memory and synchronise with the server upon establishing connexion [13,36].

3.4.3. Affiliation of principal investigator and mobile data collection

Data managers were more likely to use mobile devices when the principal investigator was not from their site. This finding raises two hypotheses. First, principal investigators from developed countries encourage data managers to use electronic data capture modalities and, because their studies usually have enough funds for the additional costs. In our view, although a study conducted with mobile data collection method will end up being cheaper than manual data

collection modality [10], the initial cost of the required hardware and skills can be a challenge [30]. Hence, we do not reject this first hypothesis. Researcher that can gather enough fund to purchase 20 or more tablets or phones for the start of a study are usually from developed countries.

Secondly, mobile data collection allows easier data access management. For example, with ODK aggregate, a data manager can grant a read-only access to the data to a remote team member. The data manager does not even have to reside at the study site as data is usually uploaded remotely. He/she can access and download the data he/she needs at any point in time, from anywhere provided that he is granted such level of access by the manager.

These are possible motivations for which investigators will communicate to data managers in opting for mobile data collection methods. When the principal investigator is from the site, he may not necessarily need these features.

3.4.4. Data entry software and platforms

Various platforms have been used for data entry. In-house platforms (12/43) and Microsoft Access (7/43) were the most used platforms. This means that researchers are not predominantly opting for open-source or freeware solutions. Epidata, CSPro, etc., are rarely used although they are entirely free. Consequently, we can assume that the cost of the platform is actually not a challenge for them.

REDCap and OpenClinica have been reported by some data managers as platforms of choice. These platforms are web-based tools, thus may often face connectivity issues. But they are suitable for health research. This is the reason why we will focus on REDCap in the development component of this research. We intend to enhance the use of web-based electronic data capture systems, by providing an open source offline mobile module based on ODK collect source code.

3.4.5. Limitations of the study

This assessment entails some limitations due to challenges we faced during the course of this research. Firstly, the coverage of our potential participating sites is 29 out of 51 (56.9%). Almost half of the sites did not respond, and the general characteristics of the sites which did not reply were not known, so we could not eventually check why they did not respond. However, with regard to the coverage of the sample, past studies indicated that internet-based data collection like what we did, rarely exceed the rate we reached; the average response rate is usually around 32.5% [45,46].

Secondly, due to small number of studies involved in this survey, we could not perform a robust multivariate regression analysis to identify factors associated with the use of mobile data collection. We therefore presented descriptive statistics and bivariate analysis from which future analysis on the topic can inspire.

Despite these limitations, our research contributes in improving availability of evidence on the use of mobile data collection tools and methods in low-income countries research centres.

CHAPTER 4: OFFLINE MOBILE DATA CAPTURE MODULE RESEARCH METHODOLOGY

In this chapter, we describe the design and scope of this research project. We also give an overview of the proposed module along with how we contrast it with previous work and the challenges. Ethical considerations are also reported in this chapter.

4.1. PROJECT DESIGN

One of the components of our work was to implement a mobile application that links with REDCap database. The main idea is to prove, by linking the module to REDCap, that the application can work across platforms. Although it is not a REDCap mobile application, it can link to its database server and upload data. Likewise, it can be linked with many other open-source EDC systems in future works.

Our mobile application is designed for offline health research mobile data collection. The application is able to asynchronously link to the web-based EDC system and load data upon establishing internet connectivity.

Our requirements analysis was through literature review and the survey we conducted in INDEPTH affiliated HDSSs. REDCap was the most used EDC system in HDSSs according to the results of the survey. In addition, unreliable internet connectivity was one of the major barriers to the use of mobile devices for health research data collection. Therefore, proposing a mobile application that collect data offline and synchronise with the database server upon establishing internet connexion, can contribute to the utilisation of mobile devices in health research data collection.

We adopted ODK-Collect as the offline Application Programming Interface (API) because of its modular developmental approach. We reviewed its source code and identified the optimal approach of implementing the application in Android.

The initial ODK-Collect has five entries on the main menu as shown in Figure 4. This is the screen that ODK-collect pops when it is started. Those five entries allow users to start filling forms by clicking on the first button (“Fill blank form”) or to edit forms that have already been filled and saved in memory (second button from the top). The third button from the top (“Send finalized form”) is for sending filled forms to servers. The last two buttons at the bottom are for getting new forms from a server and deleting forms that are already been saved in memory if the user does not need them anymore (Figure 4).

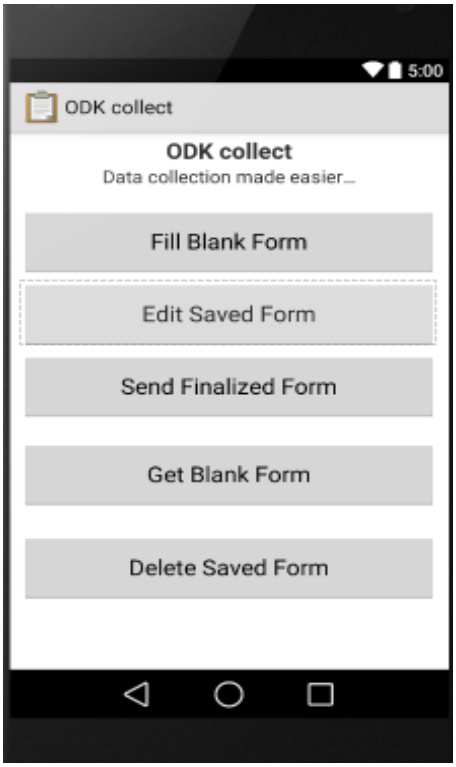

Figure 4: ODK Collect main menu

Our approach was to provide a sixth entry through which user can (1) access the internet, (2) download a form from their EDC system and (3) convert it into XForm. After these tasks, our module uses ODK-Collect functionalities to collect data and store it in the phone’s memory. Finally the last feature of CTLS comes in to synchronise the phone with REDCap database server. Figure 5 below shows the different activities and the order in which the tasks are performed. Rectangles represents what is commonly called “activities” in Android

programming. They are composed of layout with a screen displayed on the phone or tablet and Java code that execute a specific task on clicks. Diamonds represent processes that do not necessarily show a screen but execute task in the background.

The main menu on the top right rectangle of Figure 5 is the starting point. There is an entry button that does not exist in ODK (shown on the left): “get CTLS form” button. From that button the chart indicates, with a row down, to a window that will pop and ask whether the user wants to go to REDCap website or to use an existing REDCap file stored within the phone or tablet. Either way, he will get the form and parse it (bottom left rectangle). After confirmation of the parser, the user is taken back to the main menu and can get the form, and fill it using the other buttons of the main menu. Then collected data can be loaded to the database server (top right rectangle). This uses the server details saved in the preferences options of the application (server name, database name, user and password). All these steps are indicated in Figure 5.

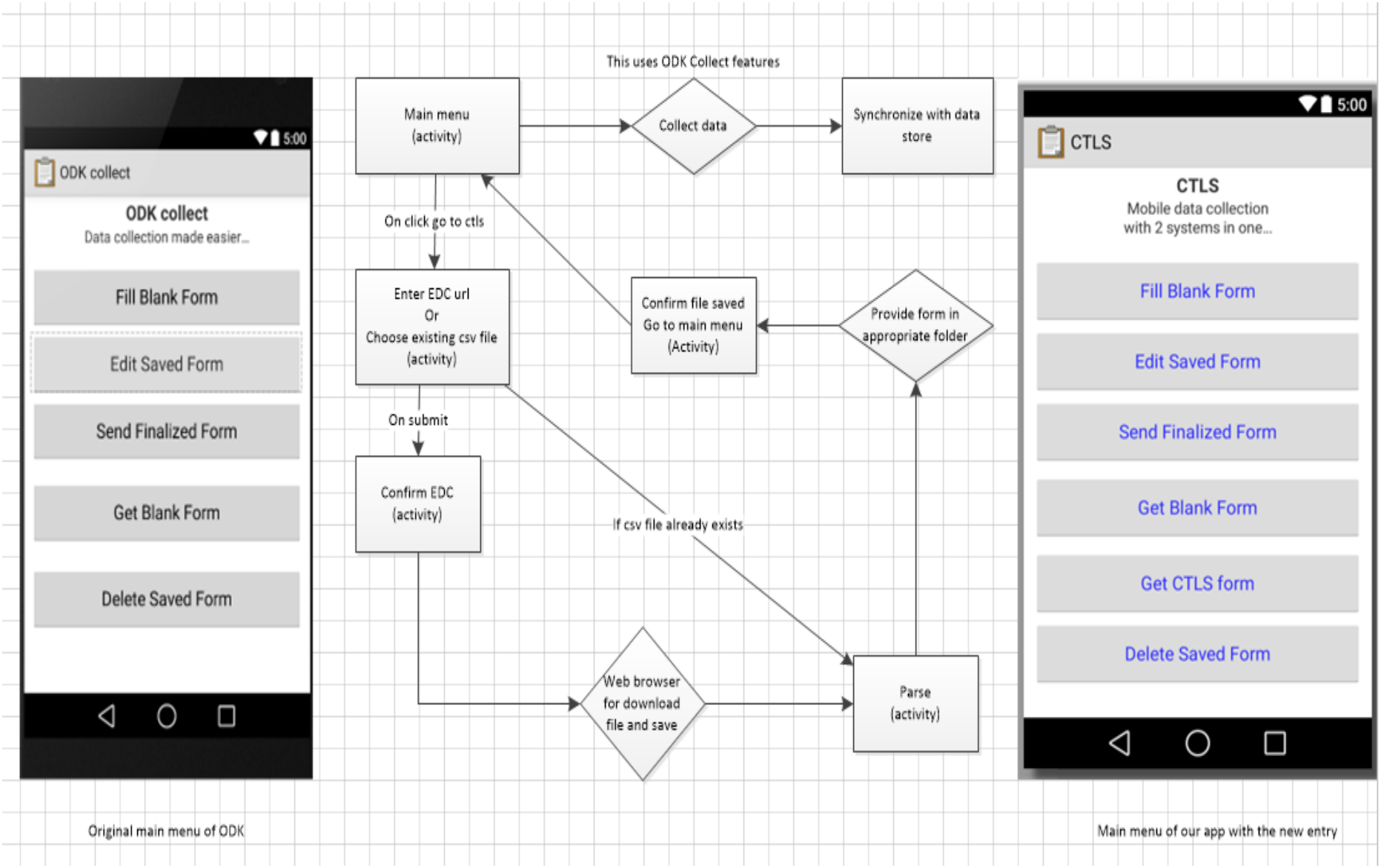

Figure 5: Flowchart of activities in CTLS

All the development tasks was done using Java 8 and Android API 5.1 in Eclipse Kepler integrated development environment (IDE).

4.2. SCOPE OF THE RESEARCH PROJECT

In this project, we developed an application (CTLS) to address a specific demand for mobile data collection at INDEPTH network affiliated HDSSs. We worked with existing open-source tool, namely ODK Collect source code, rather than building a completely new electronic data capture system. Users can collect data with our application while using REDCap database server, dashboard, and other REDCap utilities. They can use either desktop data entry or mobile application or even both simultaneously. Regardless of whether they use mobile or desktop application, all data are synchronised in some form in a data repository. This work will allow users to benefit from both CTLS offline data collection capability and REDCap EDC systems functionalities, such as logging and audit trial, visit scheduling etc., especially for clinical trials and longitudinal studies.

The requirements we worked to meet during the implementation were that the application should be able to:

- Run on Android tablets and smartphones;
- Download CSV file from REDCap server and convert it into XForm stored on the mobile device. The XForm can also be sent or copied from the appropriate folder in the phone memory;
- Display the stored XForms for offline data capture;
- Store completed forms data in the device memory
- Send the filled forms data to REDCap server using secured Wi-Fi or cellular internet connexion asynchronously.

4.3. OVERVIEW OF THE PROPOSED OFFLINE DATA CAPTURE MODULE

Our module leverages ODK-collect by adding functionalities to generate XForms, collect data and link to REDCap database, which is suitable for longitudinal studies and clinical trials. The components of our module and how it works are presented in Figure 6. CTLS offers 3 main functionalities: form provider, offline data collection and storage and asynchronous data loading into REDCap database.

The starting point of Figure 6 is the orange circle for the start of CTLS. Circles denote the functionalities, and rectangle the tasks that go with it. For instance, the form provision can be achieved only by first saving the CSV file, converting it and then saving the XForm for it to be displayed when it is needed. The next functionality is offline data collection. It is to display form, allow users to fill them and save data in memory. The metadata needed in this project is the data dictionary stored by REDCap (Figure 6).

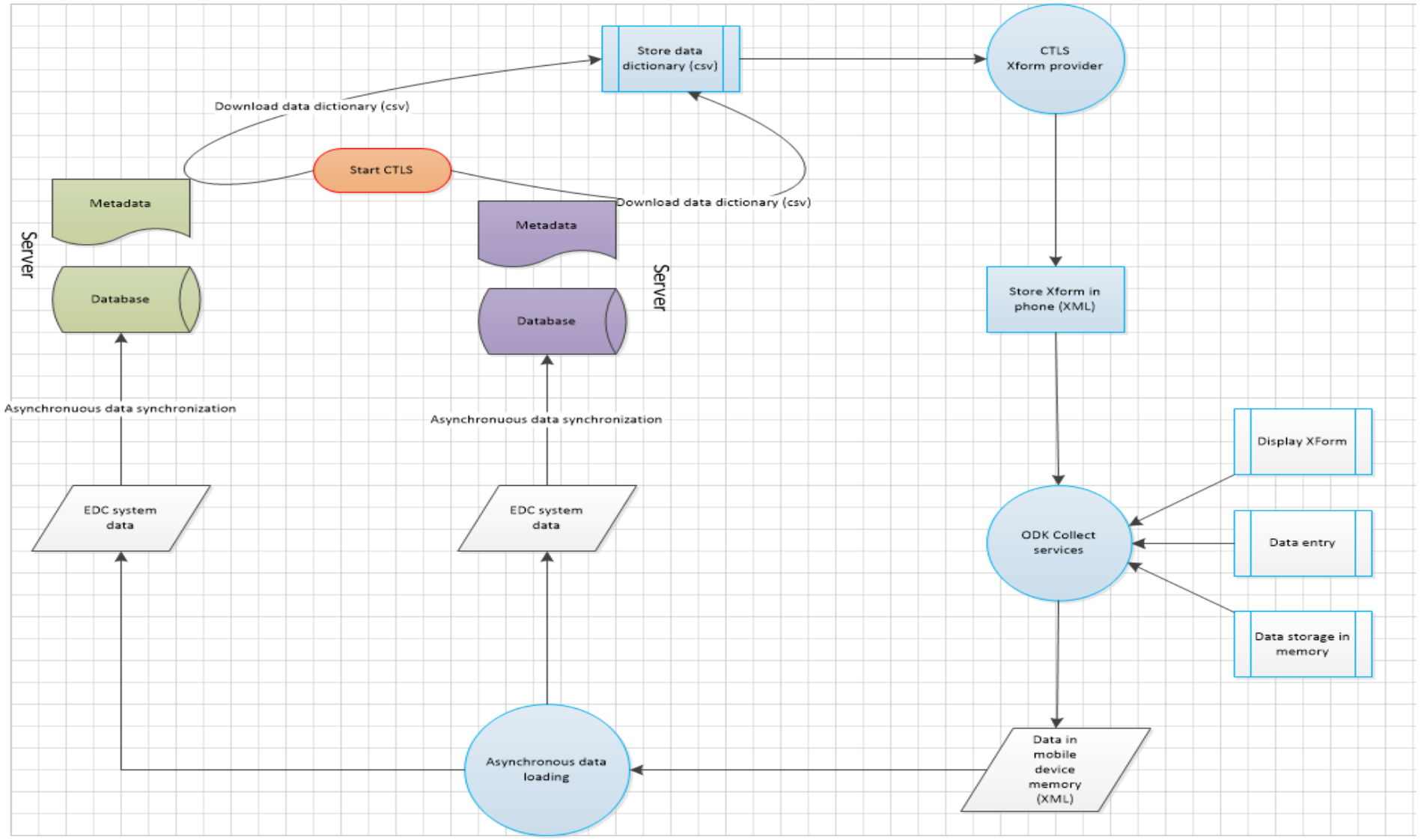

Figure 6: Architecture of CTLS

4.3.1. Form provider

The principle of the CTLS XForms provider is to download the data dictionary from REDCap server and convert it into an XML form that is compatible with either CTLS or ODK Collect. These XForms have to meet certain requirements (XML format, namespace are specific and cannot be changed, should contain head and body with specific parts that should go into each of them, etc.) and be in a very specific format for it to be recognised and work with CTLS as well as ODK Collect. Therefore, CTLS form provision from REDCap implies two steps: fetching the data dictionary from the server and converting it from CSV format to an XML XForm format.

- ***Download REDCap data dictionary***

CTLS has been developed in such a way that a button, captioned “Get CTLS form”, that we added on the main menu gives access to REDCap metadata. Users can log in using their REDCap credentials to go to get specific resource. The aim of this connexion is to download the data dictionary stored by REDCap in a CSV format. It is assumed that the project associated with this data dictionary has already been created in REDCap.

The CSV data dictionary is stored in a folder called “longitudinal” within CTLS root folder in the device (see screenshot below). We added this folder in CTLS for the purpose of storing CSV files prior to converting them into XForms. This folder is generated during the installation of the application on the device. Figure 7 shows screenshots of CTLS main menu and the folders.

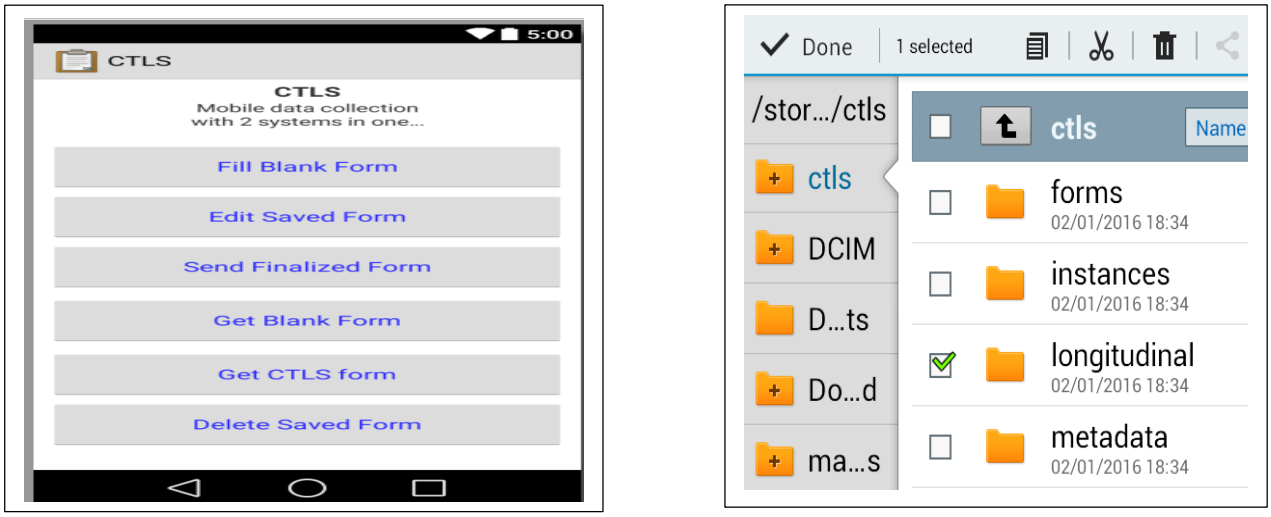

Figure 7: CTLS main menu and root folder

Alternatively, users can download the CSV file (data dictionary) right from REDCap web link and copy it into the folder called "longitudinal" via a USB cable.

At the end of the download process, the application takes the user to another interface where he/she has the option to convert files from "longitudinal". In Figure 8, we display an overview of REDCap data dictionary. This is a CSV file opened with Excel. In column A the dictionary stores the variables names, followed by column B that stores the name of the form (either exit or entry form in our project). Data types and labels are in column D and E respectively (Figure 8).

	A	B	C	D	
1	Variable / Field Name	Form Name	Section Header	Field Type	Field Label
2	record_id	entry_form_nairobi		text	Record ID
3	start_time	entry_form_nairobi		text	Start time
4	fwcode	entry_form_nairobi		text	Field workers code
5	interview_date	entry_form_nairobi		text	Date of the interview
6	name	entry_form_nairobi		text	Name of the person who has changed residence
7	household_id	entry_form_nairobi		text	Current household ID
8	relationship_code	entry_form_nairobi		text	Person's relationship to the current household head
9	room_id	entry_form_nairobi		text	Id of the room where the person currently sleeps
10	individual_id	entry_form_nairobi		text	ID of the person who has moved
11	rtesult	entry_form_nairobi		text	Result of the interview
12	respondentmoved	entry_form_nairobi	Respondent's particulars	yesno	Is the respondent the person who moved in?
13	full_name	entry_form_nairobi		text	What is your full name

Figure 8: Overview of Nairobi exit and entry forms REDCap data dictionary[7]

[7] Opened with Microsoft Excel for readability

- ***Converting REDCap data dictionary into XML XForm***

To convert the CSV format data dictionary into XForm, we used two libraries openCSV 2.4 and jdom. These libraries allowed us in Java, to read the CSV data dictionary and then convert it into an XML file while matching it to Javarosa XForm requirements. Indeed, our XML file had to meet some requirements specific to XForms in terms of structure and namespaces. Hence, the elements of the data dictionary have been fetched and written to the XML file at their specific places.

The required XForm is organised in different parts:

- Namespace

Namespaces are specific to the type of XML file we are dealing with.

- Head

The head is made up of the tittle and the model. The latter is composed of instances and bindings. Instances contain data attached to the specific variables while bindings handle data types, check controls and constraints.

- Body

The body contains all the information needed to display the form as it has been customised.

The screenshot below shows an XML format overview of Nairobi Urban HDSS entry form and how it is displayed on CTLS (Figure 9).

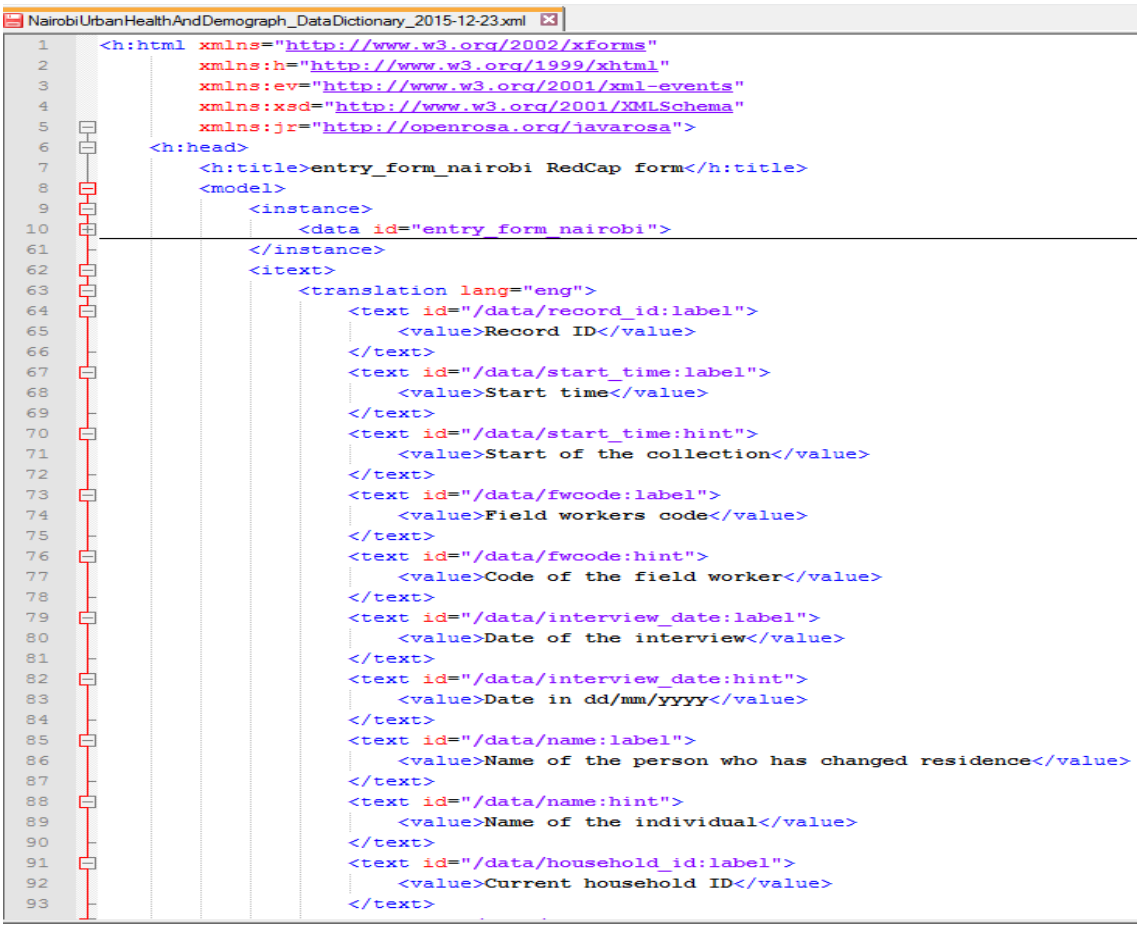

Figure 9: Overview of XForm from Nairobi Urban HDSS entry form

The final XForm obtained after converting the CSV file is then stored in ctls/forms folder. From that folder, one can access forms by going to the application main menu, clicking on “Fill blank form”, and selecting the appropriate form to start entering data.

4.3.2. Offline data collection and storage

This functionality of the module is to display forms and allow the user to complete them. And then, the collected data will be stored in the memory of the device.

As long as the CSV file is converted and stored in the ctls/forms folder, CTLS can access it and collect data offline. It can be edited, set as finalised and sent to the database. The second and third button on the main menu give those options (see Figure 7).

In the memory of the device, data is stored in XML format. The application gives the options of setting it to either sending data automatically, or temporarily storing it for a supervisor to check before sending. In the first option, as soon as the form is marked as finalised, data is sent to the database. In the second option, it will not be sent until it is checked and validated.

4.3.3. Asynchronous data upload

The application has to communicate with REDCap server via internet connectivity. To achieve that, we provide the database server details (host name, database name, username and password) as well as the REDCap project name in the preferences option of the application.

To send data, CTLS reads all the XML data stored in its memory, and loads them in the database provided that the credentials match.

REDCap uses MySQL database server, and the connexion with CTLS is performed through a web-service that uses HTTP client and PHP files that we wrote. In our PHP code, we embedded Standardised Query Language (SQL) queries to update the table in which we wish to send data. Three PHP, 3 files are called in sequence. First, a file with an embedded query fetches the project_id, and then, the second fetches the event_id from the REDCap database tables called redcap_projects and redcap_events_forms. Finally, the last file updates the table redcap_data by inserting the 5 elements of each row.

For this synchronization, CTLS uses a feature in android called *asynTask activity*, which allows to programme tasks that we want to be performed asynchronously, in the background while the device is performing other tasks, without any action or interruption under certain conditions. Our programme (activity) responsible for connecting to the REDCap database is implemented using asynTask. The web-service is embedded in that activity. And in the web-service, there is an embedded code that calls the PHP files in the order specified above, and performs the tasks in the background, upon establishing internet connectivity.

We implemented our module to be consistent with the design of redcap_data table. Each and every value (not row) entered in a field of a form in REDCap is stored in a separate redcap_data row along with 4 other variables: the project_id, event_id, record, and field name (Figure 10). But when the user enters data using the CTLS, he/she provides the field "value" required by the field. Hence, through our web-service, with embedded SQL query, we first fetch the project ID and the event ID from their tables of origin in the database, respectively, "redcap_projects" and "redcap_events_forms". This gives the 2 first values of the row. And then, we get the 3 last values from the data in the XML data file from CTLS memory. For each record, CTLS stores data on XML file in its memory along with the form name and variable names (which are nodes of the XML file).

To upload data into REDCap, we read the XML file from the device memory, and extract the values. We also fetch the form name (which allows us to get the corresponding event_id) and the nodes (which give values for the field_name column). This gets all the 5 values needed for the row: project_id (from preferences), event_id, record (from XML first node), field_name (from the node) and value which is the actual data entered in the fields.

This entire process occurs in the background using Android "asyncTask". Data can then be visualised or exported from REDCap desktop dashboard.

The screenshot in Figure 10 shows the structure of the table redcap_data. In this table, we can identify the 5 columns with some data added for test purposes.

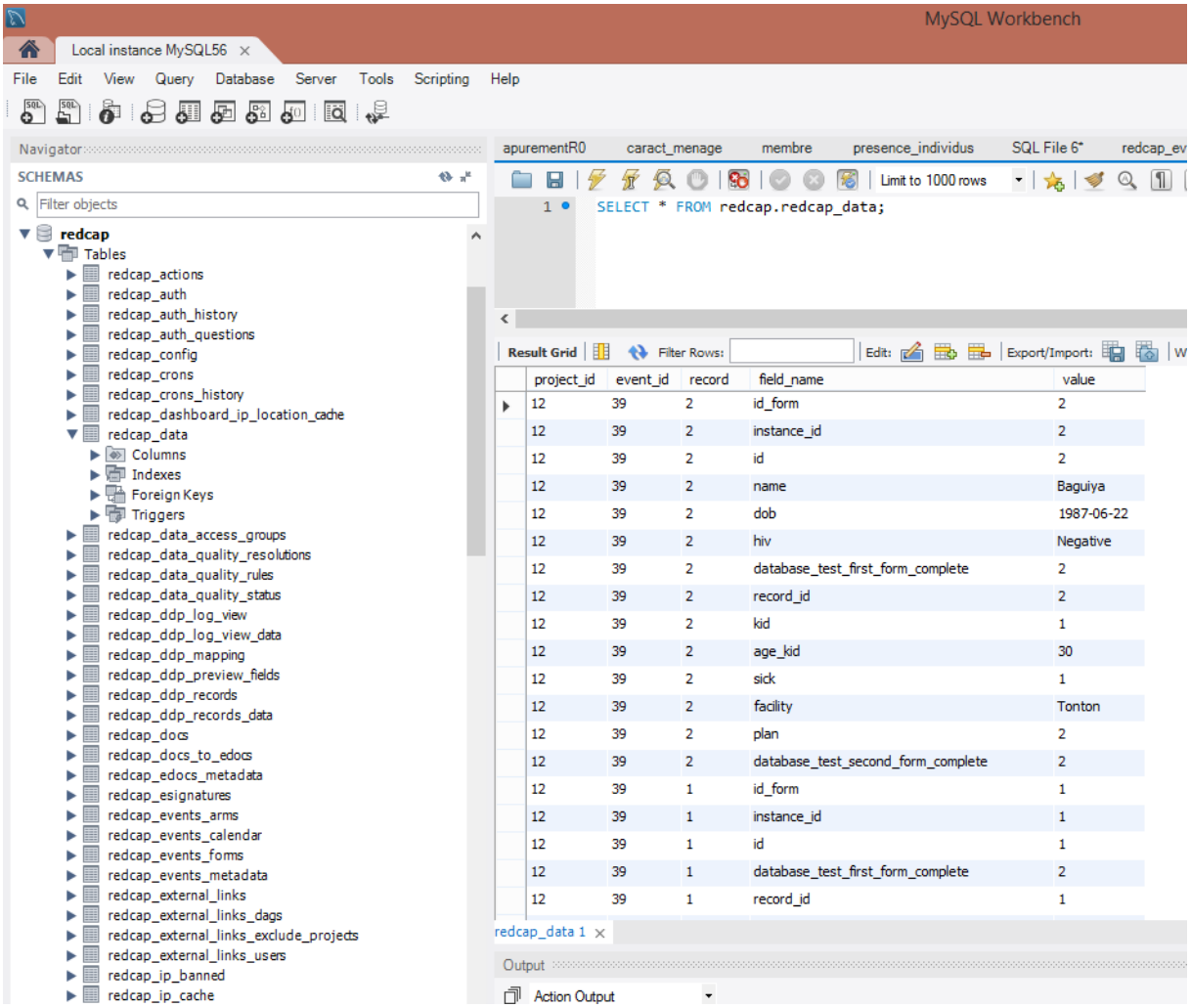

Figure 10: Overview of redcap_data table

4.4. CONTRAST WITH PREVIOUS WORK

Unlike previous contributions to the landscape of mobile data collection, our application can work across platforms: although we worked only with REDCap for now, our design allows other EDC systems, particularly OpenClinica, to be incorporated in a future work.

Although REDCap recently developed a mobile application, our work brings a possibility of using REDCap database and its dashboard with an open-source mobile application, while meeting the demand of offline data collection and asynchronous upload to the database. Finally, we built CTLS using existing tools, and unlike REDCap mobile application, ours works with both phones and tablets and it is open source.

4.5. ETHICS, DATA ACCESS AND SECURITY

In this research project, we first conducted a survey in all the HDSSs affiliated to INDEPTH network using REDCap online questionnaire. For the survey, we requested support from INDEPTH network research capacity building manager, who issued a support letter to all the site leaders. In addition to that, we sought permission from each site leader, prior to the inclusion of their site in the survey. The questionnaire targeted data managers by circulating an information sheet and a consent form. It was anonymous in terms of the respondent details. We did not need to know the identity of the respondent as all the information was about the site. No sensitive data was collected. Moreover, the collected information are kept confidential and used only for the purpose of this research project.

With regard to the development component of this work, particularly for the purpose of the testing, we requested access to forms and data collected in the Nairobi Urban HDSS. We used the Entry and Exit forms for 100 individuals in the HDSS. To ensure confidentiality, we performed all these tests using Wi-Fi network of APHRC. We setup a separate database on a test server to avoid interaction with APHRC live database.

The protocol of this study was submitted to and approved by the University of the Witwatersrand Faculty of Health Sciences Human Research Ethics Committee under the clearance n° M141173 on May 3rd, 2015.

4.6. CHALLENGES AND LIMITATIONS

EDC systems data models and dictionaries are stored in different templates and formats. For instance, REDCap data dictionary is a CSV file with a specific template. This module will only work for the EDC system we have incorporated so far, namely REDCap. For the sake of this project, we only interfaced with REDCap. We are currently designing the module to be scalable. Future work will focus on how to add more EDC systems like OpenClinica.

In addition, a REDCap data dictionary is only generated from the metadata entered while creating the project and its forms. Therefore, exceptions can be thrown by our module should some special characters be found in the data dictionary. For instance, an escape or next line character, when included in the field labels or notes of the data dictionary, will change the format of the CSV file, and therefore make the module fail to parse it. We hence recommend special to avoid unnecessary characters in the design of the form in REDCap.

Finally, we still have to work on data security issue, since we are dealing with health research data and protected health information.

CHAPTER 5: EXPERIMENTAL SETUP AND VALIDATION

In this chapter, we describe tools and technologies that have been involved in this project, and how we setup the system to perform our tests.

5.1. EXPERIMENTAL SETUP

5.1.1. Development, tools and technologies

- *Client side application*

We used ODK collect version 1.4.5 release 1048 source code. It has been cloned from https://google.code.com/p/collect in July 2015. Development was performed using Eclipse Kepler with Android Software Development Kit (SDK) and the Android Development Kit (ADT) plug-ins.

Android API 22 was our target SDK, and the minimum SDK was Android API 10. This means that our module is guaranteed to work on the release of Android 5.1 API 22, and can be supported by mobile devices running previous version of Android from API 10. Those running Android API 9 for instance, will not be supported.

We have used the Java packages OpenCSV 2.4 for the parse component of the module. The web-service used HTTP client and PHP code saved on PHP files.

- *Server side environment setup*

We have setup a REDCap server on a Windows machine running Windows 8.1. REDCap was provided by the University of the Witwatersrand REDCap administration. We used REDCap version 6.1.3. The database server was with MySQL 5.6, as REDCap works with MySQL server, which we installed along with Apache24 server.

5.1.2. Testing and prototyping

CTLS was tested with a Samsung Galaxy tab 3 (7 inches) running Android 4.1 throughout the development process.

For the purpose of testing the application, we designed, on desktop, REDCap questionnaires for the exit and entry questionnaires of the Nairobi Urban HDSS. Then, we used CTLS to download the forms and convert them to get XForms. Using those forms, we entered 50 records of each forms that were already entered in NUHDSS database.

We tested the behaviour of the module by turning on and off the Wi-Fi, and see how data was synchronised with the database server and loaded. Through this process, we tested the module regarding data integrity.

5.1.2.1. Form design

The questionnaires have been designed using the desktop online form builder of REDCap. REDCap has different approaches in designing forms on the interface: cross-sectional surveys and longitudinal studies. We designed our forms within a longitudinal study project. This is because it is more complicated: a tool that can handle longitudinal studies can easily take care of survey, but not vice-versa.

In REDCap, creating the project and forms allow us to get the database in MySQL as described in the section above. Those forms are the ones to be fetched by CTLS and displayed onto the screen for the user. Screenshots of the forms are shown below (Figure 11). Figure 11 presents the list of forms on dashboard of REDCap. We can see that the entry form entitled “Entry Form Nairobi” has 26 fields and the exit form entitled “Exit Form Nairobi” has 20 fields. This view allows user to get access to each form by clicking on it (Figure 11). This is how the main menu displays forms in the longitudinal study project in REDCap, as belonging to the same project, unlike in cross-sectional surveys.

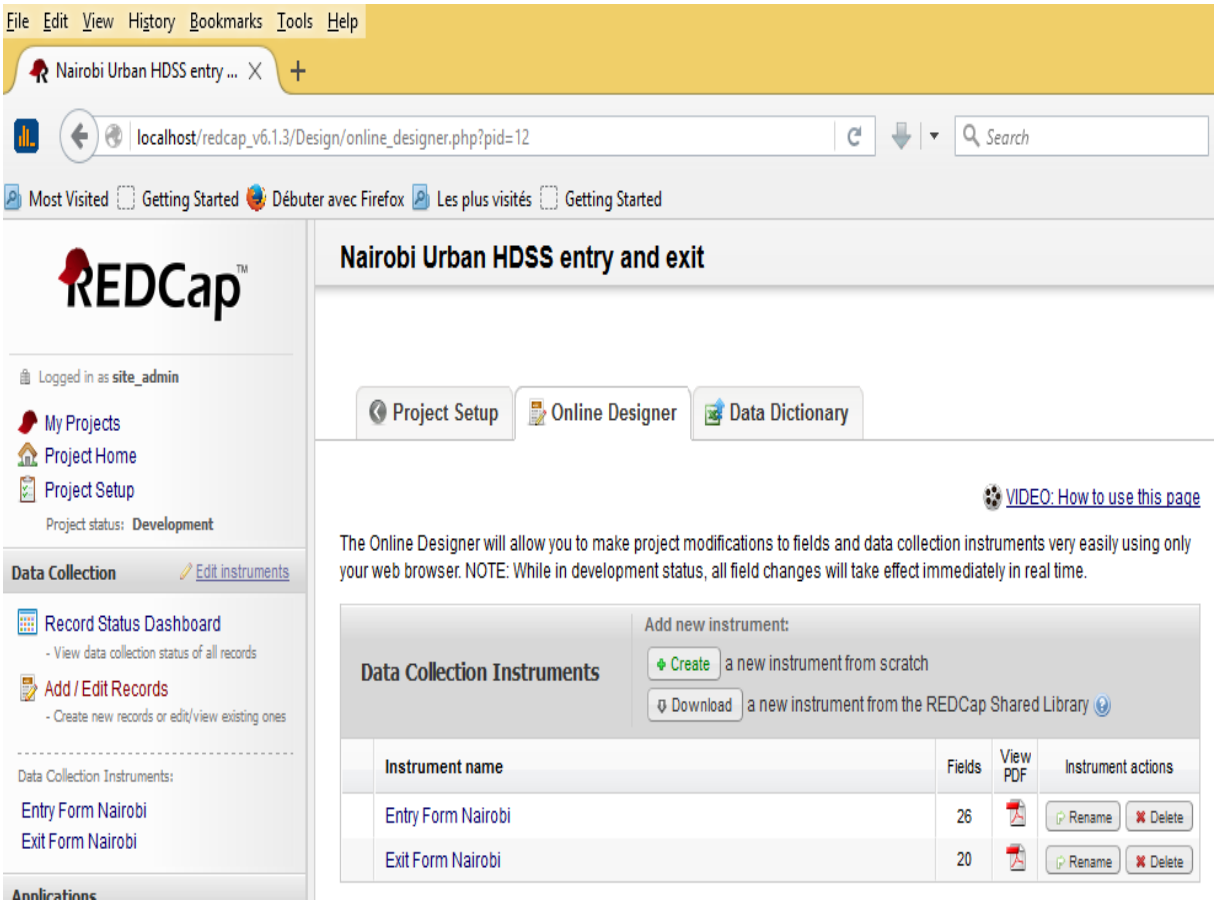

Figure 11: Forms on REDCap project interface

In Figure 12, the top 4 questions of each of the forms as displayed by REDCap for data entry are presented.

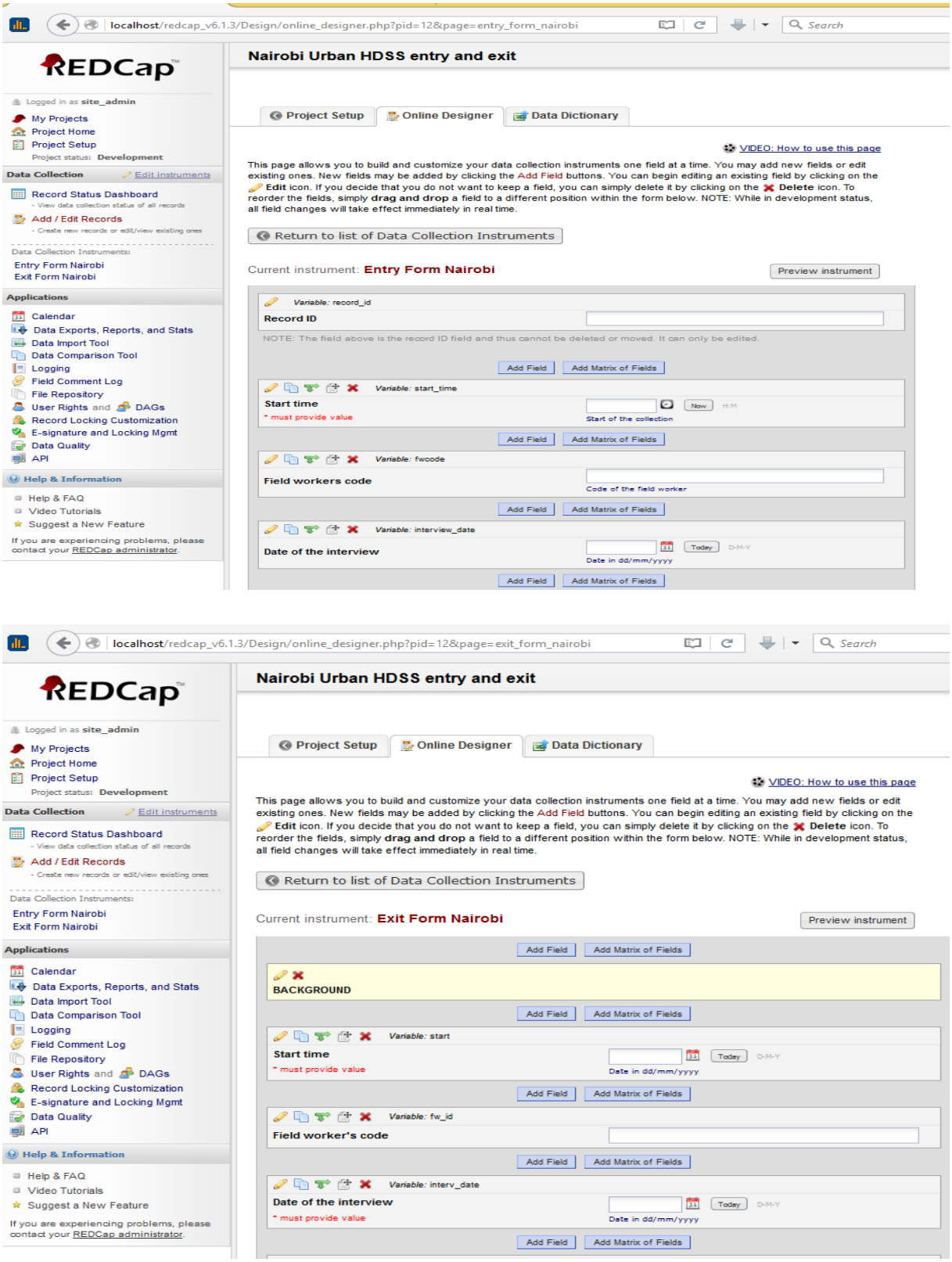

Figure 12: Forms as designed in REDCap

Figure 13 shows how fields are displayed with CTLS. The first screenshot shows a text field in which the user is expected to enter the relationship between the head of household and the respondent. The keypad is used to type the answer. The second screenshot shows a date variable. In this case, CTLS displays a calendar; the user just select the appropriate date (Figure 13).

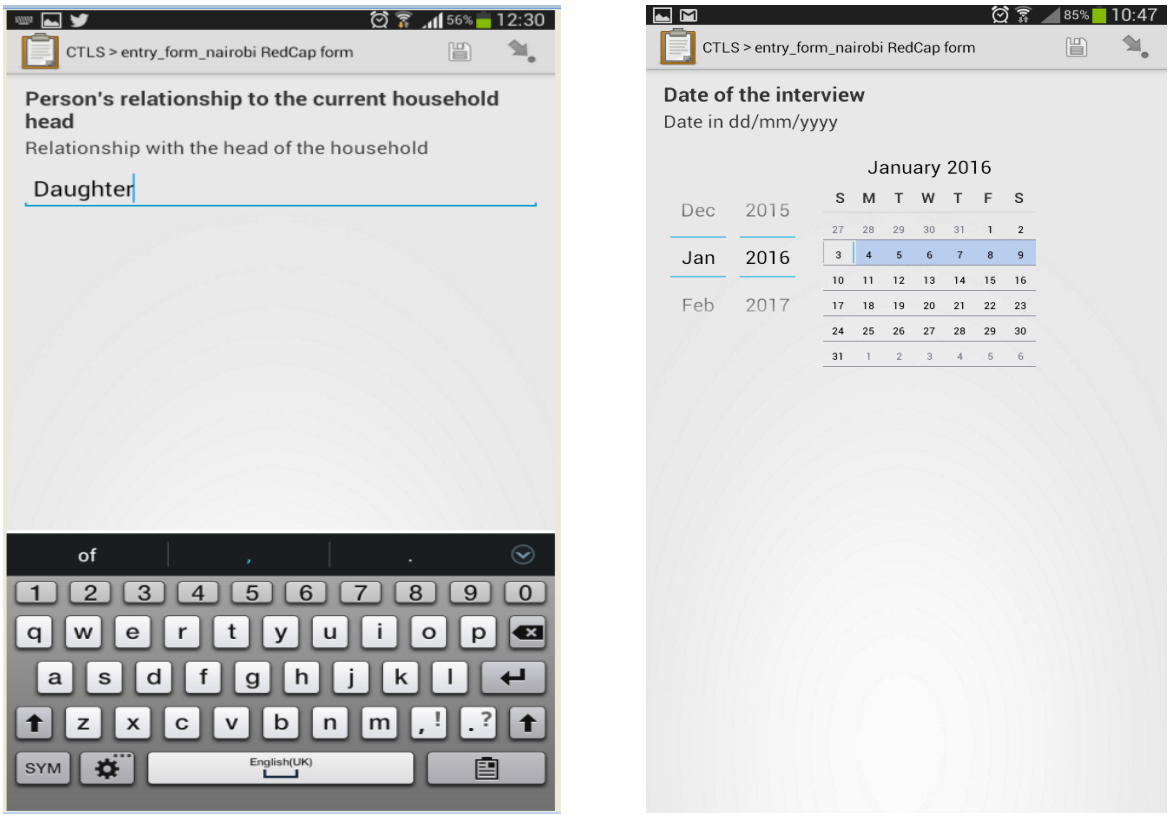

Figure 13: Example of CTLS text and date field

5.1.2.2. Data synchronisation

We installed PHP, Apache, MySQL and REDCap and then, we entered the Internet Protocol (IP) address of the machine as well as all credentials to access the database server in the preference of CTLS. The PHP files have also been copied into the REDCap server configuration folder. This establishes the connexion with the database and the data was then uploaded.

During these tests, we assessed the atomicity and data integrity with REDCap database:

- Atomicity

A batch of records is sent to the database as a unique transaction. Failure due to interruption of internet connectivity or when we turn the tablet off while the data was being sent, cause the entire process to fail and the records remain in the tablet memory if the transaction is not completed. Thus, either the form datasets are sent in its entirety or not at all.

- Data integrity

We paid specific attention to exceptions thrown by REDCap database. The aspect we dealt with was that, because the field "record" of the table redcap_data stores a value entered on the tablet for the first field of each REDCap form, different users can enter the same value for different forms. This does not actually throw an exception, but it changes data integrity and the way rows are identified in REDCap.

When CTLS converts data dictionaries, it also takes care of the check controls: all constraints in data dictionaries are translated into XForm constraints accordingly. Consequently, data allowed at the application level could not violate constraints at the database level.

Even when we had many sets of records, each particular form is parsed, and its data sent to the database as a unique transaction, one after the other. To achieve that, we used the credentials of the database that we stored in the PHP files in the application server. Each transaction is handled by a specific logging script and transaction query. Forms are handled one by one in the background.

5.2. DATA SOURCES AND DATA PROCESSING METHODS

Due to time constraint, we could not setup a study to evaluate this module on actual clinical or health related study. Instead, we used a sample of previously collected data from Nairobi Urban Health and Demographic Surveillance System entry and exit at the African Population and

Health Research Center (APHRC), Nairobi, Kenya, to re-enter data using our module in order to verify the different components and perform all our tests.

CHAPTER 6: CONCLUSION

For their core follow up, INDEPTH affiliated HDSSs mainly use manual modality for data collection. Although the emergence of mobile data collection is on the rise, the rate of utilisation in monitoring core health and demographic events is still low. This does not seem to be attributable to a lack of computational or programming skills because conversely, mobile devices are being used for embedded cohort studies. The rate of utilisation of mobile devices, mainly tablets, was much higher in embedded studies. Data quality and timeliness were the main motivations. Lack of reliable internet connectivity and initial cost of devices for mobile data collection are factors that still hinder the use of these technologies in health research data collection. The current trend of adopting OpenHDS could improve the rate of utilisation of mobile devices in HDSSs, and bring its advantages to improve and facilitate HDSSs data management.

We implemented a mobile application for embedded cross-sectional and longitudinal studies that can work across platforms. The application adds to ODK-collect functionalities, the capability of linking to a REDCap database, fetches metadata, converts data dictionary into XForm, displays it on tablets as well as smartphones for mobile data collection. Collected data can be uploaded into the REDCap database asynchronously. Finally, users can browse and view in their normal REDCap project as if data was entered using REDCap desktop. This module provides an alternate option for mobile data collection, particularly with REDCap, which has been found to be the most used EDC system in HDSSs in our study. Furthermore, this application is open source. In future work we will increase its spectrum by adding other relevant EDC platforms and improve data security. We will also explore new ways of storage data such as cloud-based data store, and new connectivity option such as Bluetooth. A user's manual will be provided for this application and the source code will be made available on GitHub.

REFERENCES

1. Chen H, Hailey D, Wang N, Yu P. A Review of Data Quality Assessment Methods for Public Health Information Systems. Int J Environ Res Public Health. 2014; 11(5):5170–207.
   Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4053886/

2. WHO (World Health Organization). Framework and Standards for Country Health Information Systems. World Heal. 2008;2nd Editio(January):72.
   Available from:
   http://www.who.int/healthmetrics/documents/hmn_framework200803.pdf

3. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inf Assoc. 2013;20(1):144–51.
   Available from: http://www.ncbi.nlm.nih.gov/pubmed/22733976

4. Choi BCK. The Past, Present, and Future of Public Health Surveillance. Scientifica. 2012;2012:1–26.

5. Society for Clinical Data Management. Good Clinical Data Management Practices Society for Clinical Data Management. April 2011. Brussels: SCDM; 2011. 502.

6. Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak. 2006;6:23.
   Available from:
   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1513201&tool=pmcentrez&rendertype=abstract

7. Betjeman TJ, Soghoian SE, Foran MP. mHealth in Sub-Saharan Africa. Int J Telemed Appl. 2013;2013:482324.

8. Brandt CA, Argraves S, Money R, Ananth G, Trocky NM, Nadkarni PM. Informatics tools to improve clinical research study implementation. Contemp Clin Trials. 2006;27(2):112–22.
   Available from: http://www.ncbi.nlm.nih.gov/pubmed/16388990

9. Ganesan M, Prashant S, Pushpa V, Janakiraman N. The Use of Mobile Phone as a Tool for Capturing Patient Data in Southern Rural Tamil Nadu , India. J Heal Informatics Dev Ctries. 2011;219–27.

10. Thriemer K, Ley B, Ame SM, Puri MK, Hashim R, Chang NY, et al. Replacing paper data collection forms with electronic data entry in the field: findings from a study of community-acquired bloodstream infections in Pemba, Zanzibar. BMC Res Notes. 2012; 5(1):113.
    Available from:
    http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3392743&tool=pmcentrez&rendertype=abstract

11. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health. 2014; 14(1):188.
Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942265&tool=pmcentrez&rendertype=abstract

12. Knipe DW, Pearson M, Borgstrø m R, Pieris R, Weerasinghe M, Priyadarshana C, et al. Challenges and opportunities of a paperless baseline survey in Sri Lanka. BMC Res Notes. 2014; 7:452.
Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4118630&tool=pmcentrez&rendertype=abstract

13. Anokwa Y, Hartung C, Brunette W, Borriello G, Lerer A. Open Source Data Collection in the Developing World. Computer. 2009;42:97–9.

14. Shao D. A Proposal of a Mobile Health Data Collection and Reporting System for the developing countries [Master]. Malmö University; 2012.

15. Asare-aboagye Y. Setting Up Electronic Data Capture Capabilities. In: Global Clinical Trials Playbook. London: Elsevier Inc.; 2012. p. 163–72.
Available from: http://dx.doi.org/10.1016/B978-0-12-415787-3.00014-X

16. Babre D. Electronic data capture - Narrowing the gap between clinical and data management. Perspect Clin Res. 2011;2(1):1–3.
Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3088950&tool=pmcentrez&rendertype=abstract

17. Prokscha S. A practical guide to clinical data management. Second. Boca Raton: Taylor & Francis Group; 2011. 254.

18. Akanbi MO, Ocheke AN, Agaba P a, Daniyam C a, Agaba EI, Okeke EN, et al. Use of Electronic Health Records in sub-Saharan Africa: Progress and challenges. J Med Trop. 2011; 14(1):1–6.
Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4167769&tool=pmcentrez&rendertype=abstract

19. Boyera S, Froumentin M, Alonso JM. Mobile Data Collection in Africa. 2012 p. 11.

20. Raja A, Tridane A, Gaffar A, Lindquist T, Pribadi K. Android and ODK based data collection framework to aid in epidemiological analysis. Online J Public Heal Informatics. 2014;5(3):e228.

21. WHO. mHealth New horizons for health through mobile technologies. Geneva; 2011 p. 112.
Available from: http://www.who.int/goe/publications/goe_mhealth_web.pdf

22. WHO. eHealth. WHO. [cited 2015 Jul 18].
Available from: http://www.who.int/topics/ehealth/en/

23. HIMSS. Electronic Health Records Standards. [cited 2015 Jul 20]. Available from: http://www.himss.org/library/ehr/?navItemNumber=13261

24. The MITRE Corporation. Electronic Health Records Overview. McLean, Virginia; 2006 p. 30. Available from: http://www.himss.org/files/HIMSSorg/content/files/Code%20180%20MITRE%20Key%20Components%20of%20an%20EHR.pdf

25. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med. 2013;106(4):130–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23564897

26. Jasmeet, Singh P. Review of smartphone operating systems like Symbian OS and Windows mobile. Int J Sci Technol Manag. 2014; 3(6):7–13. Available from: http://www.ijstm.com/images/short_pdf/1415447946_P01-07.pdf

27. Furukawa MF, King J, Patel V, Hsiao C-J, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Heal Aff Proj Hope. 2014;33(9):1672–9.

28. ITU. ICT facts and figures. 2013. Available from: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf

29. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak. 2009; 9(51). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2811102&tool=pmcentrez&rendertype=abstract

30. Giduthuri JG, Maire N, Joseph S, Kudale A, Schaetti C, Sundaram N, et al. Developing and validating a tablet version of an illness explanatory model interview for a public health survey in pune, India. PloS One. 2014 ;9(9):e107374. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4169412&tool=pmcentrez&rendertype=abstract

31. Magpi. Magpi. Magpi. [cited 2015 Jul 30]. Available from: http://home.magpi.com/resources/

32. Shirima K, Mukasa O, Schellenberg JA, Manzi F, John D, Mushi A, et al. The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. Emerg Themes Epidemiol. 2007; 4(1):5. Available from: http://www.ete-online.com/content/4/1/5/abstract

33. Kgasi MR, Kalema BM. Assessment E-health Readiness for Rural South African Areas. J Ind Intell Inf. 2014; 2(2):131–5. Available from: http://www.jiii.org/index.php?m=content&c=index&a=show&catid=38&id=100

34. Medhanyie AA, Moser A, Spigt M, Yebyo H, Little A, Dinant G, et al. Mobile health data collection at primary health care in Ethiopia: a feasible challenge. J Clin Epidemiol. 2015;68(1):80–6.

35. Seebregts CJ, Zwarenstein M, Mathews C, Fairall L, Flisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a Palm Pilot interviewing system. Int J Med Inf. 2009; 78(11):721–31.

36. Fazen LE, Chemwolo BT, Songok JJ, Ruhl LJ, Kipkoech C, Green JM, et al. AccessMRS: integrating OpenMRS with smart forms on Android. Stud Health Technol Inform. 2013; 192:866–70.

37. Sundt M. Open data kit: Aggregate Publishers Implementation Details. GitHub. 2015 [cited 2015 Aug 16].
Available from: https://github.com/opendatakit/opendatakit

38. Metts C. MDbot - Research - REDCap for the iPhone and iPad. [cited 2015 Aug 16].
Available from: http://www.drmetts.com/home

39. Institute for Clinical and Translational Science. REDCap Mobile App. 2015 [cited 2015 Aug 16].
Available from:
https://www.icts.uiowa.edu/confluence/display/REDCapDocs/REDCap+Mobile+App

40. Informa Telecoms and media. Africa Telecoms Outlook 2014. 2013.

41. Theguardian. Internet use on mobile phones in Africa predicted to increase 20-fold. 2014.
Available from: http://www.theguardian.com/world/2014/jun/05/internet-use-mobile-phones-africa-predicted-increase-20-fold

42. Shelby-James TM, Abernethy AP, McAlindon A, Currow DC. Handheld computers for data entry: high tech has its problems too. Trials. 2007;8(5).
Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1804282&tool=pmcentrez&rendertype=abstract

43. World Bank. New Country Classifications. 2015.
Available from: http://data.worldbank.org/news/new-country-classifications-2015

44. Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. Int J Epidemiol. 2012; 41(3):579–88.

45. Duncan D N. The adequacy of response rates to online and paper surveys: what can be done? Assess Eval High Educ. 2008;33(3):301–14.

46. Michael Braun Hamilton. Online Survey Response Rates and Times: Background and Guidance for Industry. 2009 p. 6.

APPENDIX ONE: SENATE PLAGIARISM POLICY

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Adama Baguiya (Student number899583) am a student

registered for the degree of Masters in Epidemiology (Research database management) in the academic year 2016.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: 

Date: February, 25th 2016

APPENDIX TWO: ETHICS CLEARANCE



R14/49 Baguiya Adama

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M141173

NAME: (Principal Investigator)	Baguiya Adama
DEPARTMENT:	School of Public Health Kenya, Nairobi
PROJECT TITLE:	An Offline Mobile Data Capture Module for Health and Demographic Surveillance System (HDSS) Studies
DATE CONSIDERED:	28/11/2014
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Gideon Nimako, Donatien Beguy and Boniface Ngayi
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	03/05/2015

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.**

Principal Investigator Signature _____ Date _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX THREE: QUESTIONNAIRE DOWNLOADED FROM REDCAP

Confidential

Mobile data collection in INDEPTH network affiliated HDSSs

This survey is carried out as part of an MSc in Epidemiology/Research Database Management (sponsored by INDEPTH network) research project at the University of Witwatersrand, Johannesburg, South Africa. The questionnaire will only take 10 to 15 minutes to complete. Thank you for accepting to participate.

You may use the mouse or the Tab button of your keyboard to move from a question to another.

Whenever you need to exit the questionnaire and resume later, please make sure you copy the return code that will then be prompted. Otherwise, you will not be able to access your survey.

For any need of assistance, please contact Adama Baguiya, abaguiya@gmail.com (+254738466899/+22670878370)

General information about the site

1 Site unique code
_____
(This code is the iShare unique code for HDSSs. Please contact us if you don't know the code of your site)

2 Name of the site
_____
(e.g. Nairobi urban HDSS)

3 Name of the country where the site is located
_____
(e.g. Kenya)

4 What method do you currently use for the routine HDSS data collection?
- ☐ Paper-based questionnaire
- ☐ Smartphones
- ☐ Tablets
- ☐ Netbooks
- ☐ PDAs
- ☐ Else

4.1 If else, please specify.
_____

5 Which mobile application do you currently use for mobile data collection for your core HDSS data?
_____
(e.g. ODK collect)

6 Which data entry software or platform do you currently use for the core HDSS?
- ○ Free platform (open source or open access)
- ○ System designed by the site
- ○ Other proprietary software (not designed by the site)
- ○ Else

6.1 If else, please specify.
_____

7 Besides the core HDSS, how many studies are currently embedded in your HDSS?
_____
(Write the number (e.g. 3 for three) or 0 if none)

Now, we would like to ask you some information about the most recent embedded studies. A maximum of 5 studies will be included, should you have more than 5.

11 In which of the following categories, the first study fits the most?
- ◯ Health system research
- ◯ Maternal and child health
- ◯ HIV surveillance, prevention or treatment
- ◯ Malaria surveillance, prevention or treatment
- ◯ Tuberculosis surveillance, prevention or treatment
- ◯ Other diseases surveillance, prevention or treatment
- ◯ Clinical trial
- ◯ Methods development
- ◯ Do not know
- ◯ Else

11.1 If else, please specify _____

12 What type of study is this study (the most recent)?
- ◯ Longitudinal study
- ◯ Cross-sectional study

13 Is this a multisite study (involving at least one other research site apart from yours in the data collection)?
- ◯ Yes
- ◯ No

14 Is this a multicountry study (involving another country in data collection)?
- ◯ Yes
- ◯ No

15 Is the principal investigator from your site?
- ◯ Yes
- ◯ No

15.1 If no, which centre/site/University is the principal investigator from? _____ (e.g. London School of Hygiene and Tropical Medicine)

16 Is the study funded by your site?
- ◯ Yes, it is self-funded
- ◯ No, it is a grant funded by another organisation

17 How long is the study supposed to last?
- ◯ 12 months or less
- ◯ 1 to 2 years
- ◯ 2 to 5 years
- ◯ More than 5 years

18 Where is the study conducted?
- ◯ Rural area
- ◯ Urban area
- ◯ Both rural and urban area

19 What method are you using or planning to use for data collection in this study?
- ◯ Paper-based forms
- ◯ Mobile devices (Phones, Tablets, PDAs)
- ◯ Else

19.1 If else, please specify _____

110 In your opinion, what are the reasons why the study did not opt for mobile devices (tablets, phones, PDA) for your data collection?
- ☐ Cost of the method
- ☐ Lack of skills for set up and maintenance of the system
- ☐ Acceptability issues among the participants
- ☐ Unreliable internet connection
- ☐ Not common practice in the site
- ☐ Else

110.1 If else, please specify. _____

111 What device are you using or planning to use for mobile data collection?
- ◯ Tablets
- ◯ Smartphones
- ◯ PDAs
- ◯ Else

111.1 If else, please specify. _____

112 In your opinion what motivated the use of mobile devices for data collection in this study?
- ☐ Data quality considerations (less errors)
- ☐ It is faster than paper-based data collection
- ☐ We have skills in our centre to do it
- ☐ We used free software
- ☐ The devices were already available
- ☐ Else

112.1 If else, please specify _____

113 How was the data sent (or planned to be sent) to the server or data store?
- ☐ Internet connection (Wifi or ethernet or mobile data bundle)
- ☐ Local network
- ☐ Else

113.1 If else, please specify _____

114 Which software or platform do you (or are you planning to) use for data entry in this study?
- ○ RedCap
- ○ Epi-Info
- ○ Epi-Data
- ○ Cs-Pro
- ○ Microsoft Access
- ○ System designed by the site
- ○ Else

114.1 If else, please specify. _____

115 Any comment on this study? _____

116 Would you like to proceed to the second study?
- ○ Yes
- ○ No