

https://dspace.org/

Standard operating procedures (SOPS) for health and demographic research data quality assurance: the case of VADU HDSS site

Twumasi, Mieks Frenken Nyarko

2016

http://hdl.handle.net/10539/22382

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

STANDARD OPERATING PROCEDURES (SOPS) FOR HEALTH AND DEMOGRAPHIC RESEARCH DATA QUALITY ASSURANCE: THE CASE OF VADU HDSS SITE

Mieks Frenken Nyarko Twumasi

Student number: 709731



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 Data Management)

Supervisor: Gideon Nimako

Co-Supervisors: Dr. Sanjay Juvekar

Johannesburg, September 2016

DECLARATION

I, Mieks Frenken Nyarko Twumasi, declare that this is my own, unassisted work under the supervision of Mr. Gideon Nimako and Dr. Sanjay Juvekar. It is being submitted for the Degree of Masters of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University. I further declare that all sources cited or quoted are indicated and acknowledged by means of comprehensive list of references.



Mieks Frenken Nyarko Twumasi

March, 29th 2016.

University of the Witwatersrand, Johannesburg

ABSTRACT

The idea of data quality assurance and security control is to monitor the quality of research data generated from any research activity. This consists of a thorough collection of documentation regarding all aspects of the research. Data management procedures of health and demographic research constantly changes or emerges through the iterative processes of data collection and analysis and requires that the investigator make frequent decisions that can alter the course of the study. As a result, audit trails that provides justification for these actions will be vital for future analysis. The audit trail provides a mechanism for retroactive assessment of the conduct of the inquiry and a means to address issues related to authenticity of the research datasets.

This research seeks to develop an Information Assurance Policy and Standard Operating Procedures for Vadu Health and Demographic Surveillance System Site using ISACA/COBIT 5 family products and ISO/IEC ISMS as benchmark. The work proposes data assurance and security controls and measures for any given research project. To develop such SOP, there is a need to identify existing gaps and inconsistencies within the data management life cycle at VRHP site. This will allow us to establish the areas of focus for the SOP.

We used an interview-based approach to identify the existing gaps associated with data management life cycle at VRHP site. The study population included key members of the data management team. The study was conducted utilizing a self-administered questionnaire with structured and open ended questions. Purposive sampling method used to enrol 21 data management team members consisting of 13 Field Research Assistants, 4 Field Research Supervisors, 1 Field Coordinator, 1 Software Application Developer, 1 Head of Data Management and 1 Data Manager. Unstructured interviews were conducted to gather information on respective roles and responsibilities of the members to ensure maximum open interactions. Data gathering and analyses were done concurrently. Two themes arose from the

data: Current lapses in data collection at Vadu HDSS and current lapses in data management at Vadu HDSS. The response rate was 95.5%.

We adopted the ISACA/COBIT 5 guidelines and ISO/IEC ISMS as benchmark to develop SOPs to guide data management life cycle activities in enforcing data quality assurance. We also included some guidelines that can be used in replicating the SOP at other research institution.

ACKNOWLEDGEMENTS

With much appreciation, I acknowledge all the persons who helped and motivated me throughout my studies. Firstly, my thanks goes to the Almighty God for bringing me this far.

My second appreciation goes to INDEPTH Networks for fully funding my master's program and to the Advisory Committee at School of Public Health of University of the Witwatersrand for rendering enormous support and advice for the successful completion of the master's program especially Dr. Latifat Ibisomi, Dr. Eustasius Musenge and Mr. Gideon Nimako.

To my internal supervisor and program coordinator Mr. Gideon Nimako, for his enthusiasm, guidance and support during my research work and studies at University of the Witwatersrand; this thesis would not have been possible without your sound advice and encouragement. I say God bless you.

To my external supervisor Dr. Sanjay Juvekar, The Officer in Charge, KEM Rural Hospital and Vadu HDSS together with his wonderful team at Vadu HDSS in India for the assistance and tremendous support while conducting my field work and for allowing me conduct the research in your site. Special thanks to Mr. Tathagata Bhattacharjee, Head of the Data Management Team at Vadu HDSS for his patience, time, advice and suggestion on how to go about with the interview and assigning me to Mr. Sandeep Bhujbal, for his time and patience to assist me with all the translation from English to Marathi and back to English, and for coordinating with Mr. Bharat Chaudhari to organize participants to be part of this study; and thanks to all the management of the Vadu HDSS who participated in the study.

I also would like to express my gratitude to my immediate boss, Mr. George Adjei for being there for me, I say "Nyame Nhyira wo".

Last and most importantly, I wish to thank my director Dr. Seth Owusu-Agyei of Kintampo Health Research Center for tremendously supporting, encouraging and for the fatherly role he played in my life, I say "Paapa, Nyame Nhyira wo".

Keywords

ISACA/COBIT 5 family products and ISO/IEC ISMS, SOPs, IAP, HDSS and VDHSS

LIST OF ACRONYMS AND ABBREVIATIONS

BR	Belmont Report
CDM	Clinical Data Management
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
COBIT	Control Objectives for Information and related Technology
DSCI	Data Security Council of India
GCP	Good Clinical Practice
HIPAA	Health Insurance Portability and Accountability Act
HDSS	Health and Demographic Surveillance System
IAP	Information Assurance Policy
IEC	International Electro Technical Commission
INDEPTH	International Network for the Demographic Evaluation of Populations and their Health
ICMR	Indian Council of Medical Research
IS	Information System
ISACA	Information Systems Audit and Control Association
ISMS	Information Security Management System
ITG	Information Technology Governance
IT	Information Technology
OSI	Open Systems Interconnect
QA	Quality Assurance
SOP	Standard Operating Procedures
SOX	Sarbanes-Oxley
VRHP	Vadu Rural Health Programme
KEMHRC	King Edward Memorial Hospital Research Centre, Pune
DM	Data Manager
FC	Field Coordinator
FRAs	Field Research Assistants
FRSs	Field Research Supervisors
HDM	Head of Data Management
ITO	Information Technology Officer
SADs	Software Application Developers
FDA	Food and Drug Administration
IAP	Information Assurance Policy
ICH	International Conference for Harmonisation
DMT	Data Management Team
DMP	Data Management Plan
QC	Quality Control

Table of Contents

DECLARATION I
ABSTRACT II
ACKNOWLEDGEMENTS IV
LIST OF ACRONYMS AND ABBREVIATIONS V
CHAPTER ONE: INTRODUCTION 1
1.1. BACKGROUND 1
1.2. PROBLEM STATEMENT 2
1.3. RESEARCH MOTIVATION 3
1.4. AIM AND OBJECTIVES 3
1.5. OUTLINE 4
CHAPTER TWO: BACKGROUND AND RELATED WORKS 5
CHAPTER THREE: INFORMATION ASSURANCE STANDARDS AND PRACTICES 7
3.1. HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA), INDIAN COUNCIL OF MEDICAL RESEARCH (ICMR), AND DATA SECURITY COUNCIL OF INDIA (DSCI) 7
3.2. CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM (CDISC) 8
3.3. CONTROL OBJECTIVES FOR INFORMATION AND RELATED TECHNOLOGY 5 (COBIT 5) 9
3.4 CLINICAL DATA MANAGEMENT SYSTEMS (CDMS) 10
CHAPTER FOUR: INFORMATION ASSURANCE AND DATA QUALITY AT VADU HDSS 12
4.1. RESEARCH METHOD 12
4.1.1. Research Design 13
4.2. STUDY SETTING 15
4.3. POPULATION AND SAMPLING TECHNIQUES 16
4.4. DATA COLLECTION 18
4.5. DATA MANAGEMENT AND ANALYSIS 18
4.6. LIMITATIONS 22
4.7 ETHICAL CONSIDERATION 22
CHAPTER FIVE: FINDINGS AND DISCUSSIONS 24
5.1. HISTORY OF PARTICIPANTS 24
5.2. THEMES ARISING FROM DATA 26
5.3. DISCUSSIONS 28
CHAPTER SIX: INFORMATION ASSURANCE AND DATA QUALITY SOPs FOR VADU HDSS SITE 32
6.1 VADU HDSS DATA LIFE CYCLE 32
6.2 INFORMATION ASSURANCE POLICY 33
6.3 STANDARD OPERATING PROCEDURE (SOP) PROPOSED FOR VADU HDSS 37
6.4 RECOMMENDATIONS 42
CHAPTER SEVEN: CONCLUSION AND FUTURE DIRECTIONS 44
7.1 CONCLUSION 44

REFERENCE 47
APPENDIX ONE: SENATE PLAGIARISM POLICY 51
APPENDIX TWO: QUESTIONNAIRE/INTERVIEW GUIDE 52
APPENDIX THREE: CLEARANCE CERTIFICATE 54
APPENDIX FOUR: INFORMATION ASSURANCE POLICY FOR VADU HDSS 55
APPENDIX FIVE: STANDARD OPERATING PROCEDURES (SOPs) 68

CHAPTER ONE: INTRODUCTION

This chapter outlines and summarizes the background of this work, the problem statements, the motivation or justification of the work and the organization of the research report.

1.1. BACKGROUND

International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) is a secretariat of international network which is currently made up of 49 independent health and demographic surveillance system (HDSS) field sites located in low and middle-income countries (LMICs) in Africa, Asia and Oceania[1]. All the sites conduct research which monitors new health challenges in the surrounding communities for an extended period of time. Vadu Health and Demographic Surveillance System (Vadu HDSS) is one of the INDEPTH field sites.

Vadu HDSS main research scope focuses on monitoring activities like immigration, emigration, births, deaths, causes of deaths and socio-demographic data such as occupation, educational levels, employment status, household characteristics etc. Research data collected from the field are analyzed and the results provide information that enable policy makers to make appropriate decisions that will positively affect the community[2]. To accomplish other research goals, Vadu HDSS has formed research collaborations with regional, national and foreign bodies to conduct several clinical studies like phase II/III vaccine trials for meningitis, measles, HIB, typhoid, Rota-virus and many others.

To achieve the desired results, it is important for the centre (Vadu HDSS) to have sufficient information that will help maintain and improve standards of clinical data management (CDM)[3] procedures and apply suitable set of standards to achieve quality data through rapid improvement in procedures. Vadu HDSS just like other research centres seek to answer research questions through data collection and analysis. It is important to ensure that high data quality is

maintained in this process. The benefits of this high quality research data can help improve health care delivery and reduce inconsistency in data during processing as well as minimize duplication of records during data collection and processing [4].

These benefits can be obtained by using modern information systems components that can keep track and offer non-repudiation by incorporating audit trails in all research procedures. Audit trails keep track of clinical research activities and their respective changes throughout the research process.

Most co-operate institutions have achieved high levels of information assurance by using information systems audit and control association/control objectives for information and related technology (ISACA/COBIT) 5 family products [5] and ISO/IEC Information Security Management System (ISMS) standards and guidelines [6] as benchmarks. This research aims to develop an Information Assurance Policy (IAP) and Standard Operating Procedures for Health and Demographic research activities for the Vadu Health and Demographic Surveillance System Site using ISACA/COBIT 5 family products [5] and ISO/IEC ISMS [6] as benchmark. These standards and guidelines which stems from IAP will guide the data and information life cycle at Vadu HDSS site. They (standards and guidelines) will also help to overcome some of the challenges associated with clinical data quality and information assurance at Vadu HDSS.

1.2. PROBLEM STATEMENT

Presently, most of the HDSS sites do not have standards and guidelines to guide data management life cycle operations just like Vadu HDSS site. There are no systems that track and document data collected from the field through data management processes to data analysis. The challenges that arise from research conducted without a set of guidelines for the data and information process life cycle activities affect the integrity of the dataset. There are a number of barriers that hinder information and data quality assurance in research studies causing data anomalies like incompleteness, inaccuracy, inconsistency, missing data. Others include date of

enrolment of study participants being greater than date the study commenced (this causes data inconsistency) and two different study participants assigned the same identification number (this causes data duplication or missing data) are some examples of data anomalies. These anomalies have great potential to undermine research outcomes and affect data integrity[7-8].

Without appropriate standards and adequate guidelines, it will be quite difficult for research result to be reproducible; there will be no evidence of proof that protocols were followed while conducting the research and there will be no mechanism to track and log the activities carried out to achieve the research outcome. Due to this reason, there is a need for standardized processes and procedures for data quality assurance at HDSS sites.

1.3. RESEARCH MOTIVATION

Confidentiality and quality assurance of research data is very important to HDSS sites. Quite a sizable amount of research data is processed on daily basis. Thus, there is the need to take all accessible quality assurance measures to ensure that good quality research data is produced. There are a number of standards and guidelines that could be incorporated in the management of data to ensure data confidentiality, guarantee quality of research data, ensure data accuracy and prevent or detect protocol violations.

The study aims to adopt the ISACA/COBIT 5 family products and ISO/IEC ISMS as benchmark to design IAP and develop SOP incorporated with audit trail to improve data quality assurance at VRHP/Vadu HDSS site.

1.4. AIM AND OBJECTIVES

The main objectives of the study is to investigate the information and data quality assurance lapses in the existing system at Vadu HDSS using ISACA/COBIT 5 family product and ISO/IEC version 27001 family of ISMS as a benchmark, design an IAP to govern data management life cycle activities in Vadu HDSS using ISACA/COBIT 5 family product and

ISO/IEC version 27001 family of ISMS, develop an SOPs and guidelines that maps information assurance policy to guide the operational activities of Vadu HDSS.

1.5. OUTLINE

The rest of the research report is organized as follows: Chapter 2 discusses in details related works conducted in the past. Chapter 3 outlines the information assurance standards and practices. In chapter 4, the methodology, research design, study setting, population and sampling techniques, data collection and data analysis are discussed, followed by the limitations, ethical consideration as well as summary. Findings and discussions of the study are outlined in chapter 5. The proposed information assurance and data quality SOP for HDSS sites are discussed in details chapter 6. Finally, conclusion and future directions of the study is presented.

CHAPTER TWO: BACKGROUND AND RELATED WORKS

Previous studies have looked at the use of guidelines and standards to ensure good and high data quality which can be replicated to answer specific research questions. A number of investigators have adopted the ISACA/COBIT 5 family products and ISO/IEC ISMS standards and guidelines to achieve good results in information assurance[9]. The adoption of these standards and guidelines has provided useful tools for information technology governance (ITG). ITG framework and supporting tools allowed IT managers to bridge the gap between control requirements, technical issues and business risks.

A study was conducted at Unisys, an IT company to outline the importance of having standardized IT strategies to support global operations and align IT infrastructure with the overall company's business strategy to help comply with SOX[10]. ISACA/COBIT 5 family product standards and guidelines were adopted and implemented to provide effective IT control and ITG framework, and SOX related controls. Outcome of the study revealed that the standards and guidelines were used to develop an approach to outsource work to third parties by identifying processes and tasks within the domains of ISACA/COBIT 5 family product[10].

Clinical researches conducted over a period of five years at the Cardiology Data Network Structure (CADANS) designed as a prospective randomized, triple-blind, placebo-controlled research among 885 men with baseline coronary cine-angiography and follow-up done after 2 years. The main objective of the study was to ensure high quality data by adopting CDM, GCP, SOPs and audit trail standards and techniques into the research process. The outcome of the study indicated that there were no serious drawbacks in managing the research rather, there were a number of benefits of networked workstations linked up to central database servers like automated data status monitoring, online message facilities, shared access and monitoring. This prevented delays, protocol violation and data errors when personnel change in the study period.

This could be due to design of protocol, data management lines and database management systems [3][11][12].

The importance of SOP in clinical researches has been proven in a selected pharmaceutical institute [13]. The research did not provide a complete set of instructions on how to create an SOP rather it provided guidelines to ensure some key elements needed in the SOP to improve routine performance tasks in the imaging and other departments were maintained to achieve highest quality. In an observational study conducted to determine and compare outcomes with accepted benchmarks in trauma care at seven academic Level I trauma centres[13], a number of SOP guidelines were adopted at the beginning of the study to treat patients. Outcome data were compared with existing trauma scales and results shown that when SOP standards were employed for severely injured patients, there was an improvement in morbidity and mortality. Outcomes of earlier studies were related but there were improvements over the older researches due to experiences gained from solutions used to address certain limitations by adopting modern techniques to bridge the existing gaps in clinical research. Different data quality assurance methods and designs were adopted in different research environments to obtain similar results. Several other studies have been conducted to observe the impact of quality assurance standards and guidelines on research outcomes [14-15]. In this study, some of the quality standards were adopted at Vadu HDSS and outcome revealed high research data quality is assured.

CHAPTER THREE: INFORMATION ASSURANCE STANDARDS AND PRACTICES

Information Assurance (IA) entails protecting your information assets from damage, degradation, manipulation and misuse; and how to recover information should any of those happen[16]. ISACA outlines information assurance as the standards which ensures that information is protected from disclosure to unauthorized users (confidentiality) within the enterprise, improper modification (integrity) and non-access when required (availability). There are a number of standards and practices adopted in research and other enterprises to assure information. In this chapter, some standards and good practices on HIPAA, CDISH, COBIT 5 and CDMS shall be briefly discussed [16][(17] and how these relate to data quality assurance and standard operating procedures for public health research studies.

3.1. HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA), INDIAN COUNCIL OF MEDICAL RESEARCH (ICMR), AND DATA SECURITY COUNCIL OF INDIA (DSCI)

HIPAA is a US law designed to improve the portability and continuity of health insurance coverage in research to combat excess, deceit, and abuse in health insurance and health care delivery as well as other purposes [18][19]. HIPAA defines security standards for data acquired from healthcare delivery, which takes into account a number of factors including: Technical capabilities of record system used to maintain research data, measures cost of security, the need for training personnel, the value of audit trails in computerized record systems, and the needs and skills of healthcare providers. Research team members who maintain and/or transmit health data are required to sustain reasonable and appropriate administrative, technical, and physical safeguards to ensure the integrity and confidentiality of health data. Data should be properly

protected from threats to the security and integrity of that data, unauthorized uses and/or unauthorized disclosure. Organizations need to be fully aware of the need to devote more resources to the protection of information assets. To address the situation, a number of organizations have set up benchmarks, standards and in some cases, legal regulations on information security to help ensure an adequate level of information assurance is maintained, resources are used in the right way, and the best security practices are adopted. In this work, the medical code of ethics in the ICMR[20] in accordance with some HIPAA standards and regulations for information security were adopted. The privacy rule covers the standards that protect the privacy and right of the study participants, de-identification of persons involved in research during data sharing, and keep data confidential from unauthorized use[21]. Data Security Council of India (DSCI) deals with best practices as a practical and accurate way to enhance adherence to data security and privacy standards to demonstrate compliance HIPAA standards as well to promote data protection[22].

3.2. CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM (CDISC)

CDISC is used in planning, collection, management and analysis of clinical and field research data, and other functions tasked with the responsibility to collect, clean and ensure the data integrity. Good data management practices are essential to the success of a research because they help to ensure that the data collected is complete and accurate[23]. The objectives of good data management in research are to ensure that research database is complete, accurate and a true representation of what took place in the research and is sufficiently clean to support the statistical analysis and its interpretation. A well designed questionnaire or interview guide is fundamental to obtain accurate and complete research data. Interview guide design should begin in parallel with protocol development. To design the data collection instrument, it essential to capture or collect only needed data that will be used for analysis, avoid collecting redundant

data and ensure that there are no redundancies**.** It is also necessary to ensure that all members of the research have adequately reviewed the interview guide before they are finalised. The interview guide or source data should be easy and quick for field team to complete. The questions in the source data should be clear and unambiguous. It is important to avoid the collection of free text as it will require coding before data can be analysed, the use of yes/no checkboxes or picklists could be considered. Also questionnaires that are translated are reviewed to ensure that the questions have a consistent meaning in all languages and there should be guidelines to complete questionnaire or interview guide, this is to assist field team.

3.3. CONTROL OBJECTIVES FOR INFORMATION AND RELATED TECHNOLOGY 5 (COBIT 5)

COBIT framework is a set of guidelines used to develop policies and adopt good practices for security and controls in IT Governance (ITG). It is important to adopt a set of standards to help ensure an adequate level of security is attained and organization's resources are used efficiently. The ISMS is designed to ensure the selection of adequate and proportionate security controls that protect information assets using ISO/IEC 27001 standards and how the requirements of confidentiality, integrity and availability of its information assets are addressed in the organization and incorporate this into ISMS[27]. The standard is the basis for the management of the organization's policy and implementation of IS.

This study adopts COBIT 5 principles to design the IAP and SOP for data management life cycle activities at Vadu HDSS that provides a robust and systematic approach to ensure that the policies are used as mechanisms to implement accepted organization strategies like data management plan (DMP)[18].

3.4 CLINICAL DATA MANAGEMENT SYSTEMS (CDMS)

CDMS information assurance standards and practices involve data acquisition, extraction, processing/coding, analysis, and transmission, storage, security, privacy and data quality assurance. CDMS have incorporated other standards and guidelines from GCP, FDA to guide clinical and field research activities to ensure that data are complete and accurate. These activities include data management through data handling, record keeping, subject and data confidentiality, safety reporting, quality control as well as records, reporting and monitoring. Data created electronically should be submitted in accordance with the FDA standards[22]. CDMS's scope include: validation of databases, audit trail for corrections in database, accounting for legacy systems/databases, copies of records and record retention. Electronic and manual data must be handled to promote consistency, efficiency as well as effective data management practices at a study level. The roles and responsibilities of all members of the research team is to ensure that the following are applied to the data that will be collected: (a) described and recorded in data dictionaries; (b) standards or terminologies that are used along with their version(s) should be listed; and (c) how data is acquired, processed and stored, where it will be stored, data handling rules and data sharing or access policy and processes are described. The CDMS information assurance and practices allows data collection process to be in a standardized format in compliance with the study protocol, regulatory requirements, allowing for efficient analysis, enables exchange of data across projects and organizations, and accompanied by a completion/instruction manual. In research, the protocol defines what data should be collected; data should be collected if specified in the protocol and only data needed for analysis should be collected using forms or interview guide manually or electronically. All appropriate individuals need to be part of the process. Interdisciplinary review is necessary. Some consideration to develop CDMS include data collection outlined in the protocol, data required by the regulatory agencies, data with all users in mind, questionnaire or interview guide

must be clear and concise, avoid duplication, request minimal free text responses, data that allows for efficient computerization.

The aim is to describe the development and implementation of a comprehensive audit trail system that meets the regulatory requirements of data quality assurance, integrity and protect study participants' privacy as well as being easy to implement and maintain[25]. Audit trail system is incorporated in the development of a database which is capable of logging data access and data changes with the correct user identifier and satisfy regulatory requirements.

CHAPTER FOUR: INFORMATION ASSURANCE AND DATA QUALITY AT VADU HDSS

The research methods and design, study setting, population and sampling techniques, data collection and analysis, limitations as well as ethical considerations used in this project are outlined in this chapter. The demographic information of the participants will be presented followed by a short history of each participant and the major themes that emerged from the data. How the findings that emerged from this study are used to create the SOPs and data quality assurance guidelines for Vadu HDSS, are also outlined.

4.1. RESEARCH METHOD

A qualitative approach and descriptive design was selected for this project. The study population were made up of key members of the DMT available at Vadu HDSS when the interviews were conducted. A purposive sampling was done to interview 22 data management team (DMT) members made up of 1 Head of DMT (HDM), 1 Data Manager (DM), 2 Software Application Developers (SADs) and Information Technology Officer (ITO), 1 Field Coordinator (FC), 4 Field Research Supervisors (FRSs) and 13 Field Research Assistants (FRAs). The interviews were conducted in English, however a translator translated the interview questions and responses into Marathi and back into English because most participants could not express themselves in English.

A set of unstructured interviews were conducted to collate data which allowed the investigator to investigate and understand the work experiences of the participants and identify lapses regarding quality assurance in data management cycle. Before conducting the interviews, a copy of the general research information sheet was given to every participant to brief them about the proposed study and also to invite them to participate in the study. A copy of the informed consent form was provided to every participant to obtain verbal and written consent as well as obtain permission to participate in the study to indicate that they have not been coerced into giving consent and that consent was given freely and

voluntarily[26] and also participants were made to understand that withdrawal from the study could be made at any time if they wished. The main question was for participants to share their work experiences as DMT members with investigator where existing lapses in the current system were identified. Permission was obtained from participants to audio tape in-depth interviews for the transcription purposes. Personal notes were also made during and immediately after the interviews. The personal notes were a written account of the salient points made by study participants during the interview sessions. All interviews were conducted in the DM's office. Responses from in-depth interviews were used to answer the questions available on interview guide found in Appendix Two. Different methods of triangulation were used during interviews where same questions from or interview guide were asked to cross check the reliability and validity of responses from participants having the same roles and responsibilities[27]. This strategy is used to obtain the maximum amount of data, to verify what was heard and what was meant by participants. Care was taken by being sensitive and not forcing participants to answer questions they preferred not to[28].

4.1.1. Research Design

A qualitative design and descriptive approach was selected for the study[29].The questionnaire or interview guide used to interview participants sought to measure data quality assurance lapses associated with data management processes at Vadu HDSS in accordance with ISACA/COBIT 5 and ISO/IEC version 27001 standards. The interview sought to get information concerning five areas. The first section in Appendix Two sought to identify various roles and responsibilities in data management; this was based on ISO/IEC version 27001:2005 structure[31] which described the procedure of managing responsibility and ISACA/COBIT 5 process and guideline version APO01 which outlines IT management framework. It tried to identify data and information flow in the data management processes. The second section sought to identify concerns associated with data collection, data processing and data preparation for analysis. This was based on Appendix G Detailed Description of ISACA/COBIT 5 Enablers information[32] which described the plan and design that would be used to capture data

collection and data process life cycle. Activities in this phase referred to the identification of objectives, the planning of the information architecture and the development of SOP and IAP. In this section, we determined the various audit trail mechanisms and techniques used in the daily data management activities at Vadu HDSS. This is in compliance to standard 1201 Engagement Planning of ISACA/COBIT 5[32] and IS Audit and Assurance Guideline 2203 Performance and Supervision and Standard 2.5.1-2.5.2. This was to enable audit engagement objectives and scope of work, project plan, work program, steps performed, evidence gathered, conclusions and recommendations. This was to aid planning, performing and reviewing audit engagements because it identifies the IS audit team members who performed each audit task and their role in preparing and reviewing the documentation, records the evidences requested, supports the accuracy, completeness and validity of the work performed and provided support for the conclusions reached.

The section also sought to identify measures and relevant plans put in place to monitor follow-up activities in the field. This is in accordance with IS Audit and Assurance Guideline 2402 Follow-up Activities of ISACA/COBIT 5 Process version 1402 and 1005 Due Professional Care standards[33] which states that there must be appropriate plans to monitor follow-up processes, timing and scheduling, reporting of follow-up activities, proposed action taken to correct or prevent potential risk and follow-up by personnel.

In the third section, we identified the storage and archiving systems for data management activities for easy retrieval, safe-keeping and for reference purposes. This is in accordance with ISACA/COBIT 5 standard 2.5.1-2.5.3[34] which ensured that obtaining and managing information should be retained and properly disposed of in accordance with organizational policies and relevant laws, rules and regulations.

The next section sought to describe audit assurance in compliance with ISACA/COBIT 5, IS Audit and Assurance Guideline 2003 Personnel Independence, section 2.0 – 2.9, Conceptual Framework 2.1.0 - 2.1.6, Threats and Safeguards 2.2.1 -2.2.6 and Managing threat 2.3.0 -2.3.6[35]. The guidelines describe

potential threats and how they could be managed. Threats fall within these categories; *Self-interest and review*: this threat arise as a result of an individual performing an inappropriate evaluation. *Intimidation*: this threat arises as a result of the fact that personnel were prevented from acting with integrity and objectivity due to perceived pressures, including attempts to exercise undue influence over other personnel. *Bias*: the threat that personnel will, as a result of political, ideological, social, psychological or other convictions, take a position that was not objective. *Management participation*: this is the threats that arose as a results of personnel taking the role of management or otherwise performing management functions on behalf of the entity undergoing an audit or assurance engagement.

The last section (i.e. section F) solicits for information assurance lapses in the current system. This is in accordance with ISO/IEC 27001:2005[30], Data confidentiality clause 3.3 of ISO/IEC 27001 which ensures that information is accessible only to those authorized to have access, Integrity clause 3.8 of ISO/IEC 27001: which gives guidelines on safeguarding the accuracy and completeness of information and process methods, Availability clause 3.2 of ISO/IEC 27001[30] which ensures that authorized users have access to information and associated assets when required.

These standards sought to improve research data quality and harmonized with healthcare standards to provide interoperability among clinical research studies. Using appropriate standards an SOP was developed to guide routine data management activities at Vadu HDSS. This is a set of instructions which specified how to perform the activities to guide data management operations; from the investigator's site through data cleaning and preparation of data for analysis. With the use of audit trail, it was easier to keep track, log and monitor what was done. This could be used to track certain activities that occurred during data management processes.

4.2. STUDY SETTING

Shirdi Sai Baba Rural Hospital of Vadu Rural Health Programme (VRHP) is a branch of and managed by the KEM Hospital, Pune that provides secondary level care to the 22 villages of the Vadu Rural

Health Program with a population of 68,000 and beyond which is located in the Vadu Budruk village in Shirur Block of Pune District. This is a non-profitable organization which integrates a multi-disciplinary approach to extend quality health care to an under-privileged rural population in a tripartite partnership with the Zillha Parishad and state Government [36][37].

VRHP/Vadu HDSS seeks to bridge the gap in the health services provided in the villages and its environs. It keeps track of health issues and demographic changes in VRHP area. Research data captured are analyzed and the results enable policy makers to take suitable decisions to address emerging public health needs. For this aim to be achieved, Vadu HDSS proposes the design of policies and SOPs to guide data management life cycle activities that will ensure high research data quality retained in the data management processes which could improve health care delivery and reduce data discrepancies.

4.3. POPULATION AND SAMPLING TECHNIQUES

The target population comprises of the key members of the DMT currently at Vadu HDSS. The DMT members met the inclusion criteria of the study. Purposive sampling was used to select the sample involving 22 members of the DMT with different roles and responsibilities. The number per roles and responsibility comprises of 1 Head of Data Management (HDM), 1 Data Manager (DM), 2 Software Application Developers (SADs), 1 Field Coordinator (FC), 4 Field Research Supervisors (FRSs) and 13 (FRAs). All those purposively sampled agreed and availed themselves for the study with the exception of 1 SAD who did not give written consent to enroll in the study because she left the study area. This reduced the sample size to 21study participants.

The sample consisted of 22 study participants who were interviewed till no new information was emerging from their responses i.e. saturation [28]. Twenty-one (21) out of the 22 gave full consent and participated in the study. Out of the 22 study participants, 20 of them were residence of Vadu which is a rural setting and 2 residence of Pune; an urban setting. The basic language spoken and written by all the 22 study participants is Marathi and 5 of them speak and write both Marathi and English. All interviews were conducted at the DM's office excluding 1 of the SADs and HDM.

Interview guide was sent to HDM by email and returned by e-mail. SAD was interviewed through Skype and also questionnaire was sent to her by email due to poor network during Skype interview. She is yet to return the filled questionnaire and the informed consent form. The participants came from the same cultural and religious background and in terms of race, they are all Indians. The demographic information of the participants is presented in Table 1.0.

Table 1.0: Demographic information of the participants

CHARACTERISTICS	**n (%)**
Roles/Responsibilities	
Field Research Assistants (FRAs)	13 (59.1)
Field Research Supervisors (FRSs)	4 (18.2)
Field Coordinator (FC)	1 (4.5)
Data manager (DM)	1 (4.5)
Software Application Developer (SAD)	2 (9.1)
Information Technology Officer (ITO)	1 (4.5)
Head of Data Management (HDM)	1 (4.5)
Residing place	
Urban (Pune)	2 (9.1)
Rural (Vadu)	20 (90.9)
Language	
English and Marathi	5 (22.73)
Marathi	17 (77.27)
Race	
White Indian	21 (95.5)
Black Indian	1 (4.5)
Gender	
Female	11 (50.0)
Male	11 (50.0)
Participants Enrolled	
Yes	22 (100.0)
No	0 (0.0)
Full Consent	

Yes	21 (95.5)
No	1 (4.5)

4.4. DATA COLLECTION

Thirteen (13) unstructured individual and in-depth interviews were conducted after gathering general information from the 21 study participants. These interviews were conducted without the investigator employing any prior information, experience or opinions in a particular area. Unstructured interviews were used to determine DMT members' experience on their roles and responsibilities and existing lapses were identified by the investigator[28]. This interview technique was appropriate for the study because it was informal interview that allowed the investigator to have in-depth understanding of participant's experience regarding their roles and responsibilities and identifying existing lapses in the current system. The responses from study participants were used to answer questions on the interview found in Appendix Two. The major question was for study participants to share their work experiences with investigator. The investigator used probes and prompting questions [28] to encourage participants to expand on their experiences. The interviews were recorded by audiotape and field notes. The audio-recorded interviews were transcribed verbatim and the field notes, which are non-linguistic utterances such as a nod, smile and sigh which could not be recorded with the audiotape, were added to the transcriptions.

4.5. DATA MANAGEMENT AND ANALYSIS

Qualitative data was manually analysed thematically. The interviews were transcribed verbatim and the transcripts were crossed checked for accuracy[27-28]. The steps in the data analyses process were as follows:

Step 1: The transcribed interviews were read and re-read to get a sense of the experience.

Step 2: Responses were summarized and categorized according to themes. Responses relevant to the phenomenon being studied were identified and coded.

Step 3: The investigator identified the meaning of the participants' experiences which were transformed and aided investigator in answering the questions on the interview guide. The investigator looked for

perceptions and emotions which were expressed by the participant's description in order to come up with findings. Data analysed is presented in Table 2.0.

Table 2.0: Data Analysis

Themes	**Findings**
Section A (Roles and Responsibilities) What are the roles and responsibilities in data management life cycles?	FRAs: Responsible to conduct house to house surveys to collect and update HDSS data within defined HDSS areas. FS/FC: Responsible for planning survey process, monitoring progress, completeness and field data quality. SADs: Responsible to design database structures based on the forms and design e-forms. DM: Responsible for creating new forms, modify existing forms, design database structures based on the forms, design e-forms, data management, storage, retrieval, archival and security of data stored in electronic formats. HDM:Responsible for overall planning, designing and execution of all data management software applications and database management.
Describe the flow of data management task.	Data Collection (FRAs:) Introduction of study in Marathi. Obtain permission for enrolment and/or capture events updates. Check updating list whether household information exists or not. Lesser literate respondents pose more problems. FS called if problem exist or change of location. FS tracks/monitors field work. FS randomly re-checks data recorded with same participants for correctness. FS/FC: Beginning a round, needed laptops, questionnaires and other materials, FRAs and database made available for data collection. Training session conducted. FRAs village allocation. Daily and weekly target specified based on households. Data collection completion days specified. FRAs weekly meetings for recorded query discussion and resolution. Unknown field visits by FRAs. Query resolution. Responses randomly re-checked. Data captured are births, deaths, marriages, pregnancies, in and out migrations and others. Ensure smooth running of field work. Recheck 10% of data. Queries as a result of communication problems. FSs' remarks recorded. Data Management (SAD/ITO): Development of HDSS software on tablet. Supports DM. Network problem resolution. Resolves software application problems. Design HDSS application on tablet. Field visits to understand data collection. Knowledge in developing software applications using Java script. More data quality features on tablet. Audio recordings during interview. Household heads photo captured. Informed consent form signed electronically by finger. GPS activated to capture location of interviews. Electronic data upload to office from any location. DM: Design HDSS application for data collection. Design query and validation application for data management. Population counts per village and per site. Laptops formatting and updates

	made. Laptops reassigned to FRAs. Declaration form signed by FRAs after laptop allocation. Laptop automatic backups. Dump database re-cleaned. Export of data from FRAs laptop to DM's laptop is updated weekly. HDM: Portable database updated. ETL scripts run for data cleaning. Outstanding queries resolved and database updated. Laptops sent to DM's office for weekly backups and updates. Qualities checks run and data cleaning. Automatic backups made on local drive during updates. Events counts per village made available to FC for weekly reports generation.
Section B (**Data collection, data processing, data preparation for analysis and audit trail**)	(FRAs/FRS/FC/SAD/ITO/DM/HDM): Manual and electronic, laptop and migrating to tablets. Data Collection (FRS/FC): Monitor FRAs using ATP. No other tracking system on the laptops except manual system. FS and/or FC un-announced field visits during data capture. Better monitoring system on tablet. Back and field QC checks, un-announced visits to randomly re-check data captured with respondents for accuracy. Identified queries recorded. Query resolution. Data Management (SAD/ITO): Unknown GPS activation and automatic audio recordings during interviews. DM: Generation of quality control checks. Data files are used for different forms. Query resolution. FS and/or FC call FRAs for location if not at known location. Training sessions completed. Field monitoring. Query discussion and resolution on field weekly meeting. Household head and updating list to guide follow ups. HDM**:** The documentations available are database diagrams and quality check rules.
Section C (**Validation mechanisms)**	Data Collection (FRAs/FRS/FC): We log unto the system with assigned username and password. Data Management (SAD/ITO/DM): Automatically generated date, time, and FRA's name after logging unto the system. HDM: Pre-validations mechanism primarily involves process validations like the interview techniques and forms and their accordance with the standard HDSS practices.
Section D (**Data storage, filing and archiving**)	Data Collection(FRAs/FRS/FC): DM backup data collected and updated data every Monday. After data collection, laptops brought for weekly backups and updates. Data Management (SAD/ITO/DM): Query resolution, quality checks run, data cleaning and automatic data backup on primary database and other devices. With the tablets. FRAs can upload data captured unto server from any location. Run job to create automatic backups on local and cloud servers. Copy sent to HDM and KEM server. HDM: Data provided on request and approval of competent authorities.

What lapses you know of in data management processes that can affect data quality?	Data Collection(FRAs/FRS/FC): False information given by participants about health and socio demographic status but the policy is that any information given to us by participants is true. Household heads with lesser education create problems. Respondents don't understand why certain information required, after washing Vadu movie right information given. Difficult to understand Vadu movie because movie more in English than Marathi. Inconsistencies in data sheet and information collected. Some participants want some discounts from KEM hospital when they visit. Some respondents do not want ids written on their doors. Major problems are communication. Respondents think only Vadu benefits. Respondents think Vadu is rich due to migration from manual to electronic. Sometimes locating FRAs is problem. Problem locating keys on keyboard. Lack experience of using laptops. Difficulty in filling e-forms, if a mistake is committed you cannot save; problems resolved after a week. While filling the e-forms, if a mistake is committed you cannot save. All questions mandatory. Wife willing to provide information but husband not available or disagrees. Problems starting the laptop. Difficult to shut down laptop. Problem accessing the HDSS system. Filling e-forms is difficult. E-forms cannot be accessed due to technical problems. Problems search names with 4 characters creating problems for 3 characters' names. Functional problems switching from electronic to manual. Hardware problems. Battery problems. Permission problems. Wrong or missing date of birth or age. Household heads changing problems. Problems with dates. Data Management (SAD/ITO): No tool available for application on the tablet. DM: Name problems, date of birth either wrong or missing. Matching data difficult. No detailed documentation of the processes. HDM: Very less standard analysis is done on the datasets hence many unknown facts are hardly reported.
Section E (**Audit Assurance**)	Data Collection (FRAs/FRSs/FC): In-appropriate evaluation could be detected by logging unto the system with username and passwords. Data Management (SAD/ITO/DM): Audio recording during data capture, GPS activation on the tablet application, in-appropriate evaluations will be detected automatically on the tablets. When database manually opened, in-appropriate evaluation could be detected. Field QC and back checks, re-checking information captured by FRAs with respondent for accuracy. Some undocumented mechanisms in place to compare results with other researchers and published papers and abnormalities are reported. FSs and FC physical monitoring and re-check FRAs work. Audio recordings on tablet. Validation and manual checks on database. Field monitoring. HDM: Some undocumented mechanism. GPS activation and audit trails incorporated on tablets application.

Section F (**Information Assurance**)	Data collection(FRAs/FRSs/FC): To gain access to HDSS application, we log with assigned usernames and passwords. Data Management (SAD/ITO/DM): Encrypted passwords. Usernames and passwords assigned to each individual to log unto HDSS application. After data cleaning, data updated, and backed up on both the local server and the cloud server (Dropbox). Without user authentication it would be difficult to get a copy of another backup copy from the automatic backup copy. Processed data stored on servers accessible to only authorized personnel. Without correct answers to questions, system unable to save data entered. No blanks. Incomplete data not possible. All questions mandatory. FRAs must choose from options provided on application. Less manual data entry. Inbuilt automatic checks on HDSS application. Quality assurance and validation checks run. Weekly backup made on all available storage devices. Data is encrypted. Only authorized persons have access to data. Data can only be access based on request and approval of Juvekar Sir. HDM: Data confidentiality handled by case specific during data sharing where ids are anonymized.

4.6. LIMITATIONS

Five months allocated to this study was not enough to test or conduct extensive evaluation of the IAP and comprehensive SOP development.

Most study participants were unable to express themselves in English language. Translating from English to Marathi and back to English could distort the actual responses from study participants. The sample size was small, this was due to few key members in the DMT and saturation level.

Simulating of IAP and SOP was not done due to time limitation.

4.7 ETHICAL CONSIDERATION

Ethical approval and clearance was sought from the University of Witwatersrand Human Research Ethics Committee. Permission and ethical clearance were obtained from the allocated site (Vadu HDSS) before the research was conducted. A formal letter of approval was also obtained from the site leader.

The following measures were applied to ensure the study was conducted in an ethical manner:

- Permission to conduct the research: The research proposal was submitted to the Postgraduate Committee of the School of Public Health at University of Witwatersrand for further assessment and approval, then after it was submitted to the Human Research Ethics Committee. Ethical clearance was obtained **(***Clearance certificate number: M141164*) found in (Appendix Three)**.** Permission to conduct the study was obtained from the Officer in Charge of the Vadu HDSS through the KEM hospital and approval given.
- Informed consent: An information document was first given to the study participants and then informed consent, in writing, was obtained for participation. The participants were allowed to withdraw from the study at any time they please.
- Anonymity of participants: Anonymity simply means protecting confidentiality and privacy of the study participants. This was ensured by not using the participants' actual names during data collection and analysis.
- Confidentiality: Confidentiality is the promise of not divulging the information obtained but only using it for the intended purpose[38]. Identification numbers were used instead of actual names. The audio tape recordings were saved with a password known only to the investigator and supervisor.

CHAPTER FIVE: FINDINGS AND DISCUSSIONS

The findings, history of the participants, and themes arising from data of the study are discussed in this chapter.

5.1. HISTORY OF PARTICIPANTS

A brief history of each participant is provided below. Identification numbers were used instead of the participants' names. Participant (P) P1 to P5, P10 to P13, P14 to P17, and P18 to P20 were FRAs and FRSs and had their interview in groups of 2, 3, 4, 4 and 3 sessions respectively. The questions were translated from English to Marathi and responses from Marathi back to English by translator P7. In-depth interviews were conducted for FRAs and FRSs in English and Marathi. P6, P7, P8, P9, P21 and P22 had their interviews conducted in English.

- ***P1***. Participant 1 was a female; a FRA working on HDSS projects. She had worked on both the manual system (hardcopy or paper-based) and electronic system (laptop or tablet). She worked with the manual system since 2011; that is 2 years. After 2013, she has been working with the electronic system (laptops).
- ***P2***. Participant 2 was a male and had been working with the electronic system for a year as a FRA on the HDSS project.
- ***P3***. Participant 3 was a female, an HDSS FRA who worked on the hardcopy from 2009 to 2011 and currently working with laptops or electronic systems.
- ***P4***. Participant 4 was a female; who worked with hardcopy from 2009 – 2011 on the HDSS project. She is currently working with the electronic system.
- ***P5***. Participant 5 was a female, FRA who worked with only laptops since 2014 on HDSS.
- ***P6***. Participant 6 was a male, an ITO and SAD who had worked with Vadu HDSS since 2014 May. He developed the HDSS application on the tablets. He resolves network problems at the centre. He assists the DM when he travels in doing part of his work. He speaks and writes English and Marathi.

- ***P7***. Participant 7 was a male. He was a Junior DM at Vadu HDSS and also a staff of INDEPTH Networks working on iSHARE project. He had conducted 3 rounds of the HDSS study. He had been working on the webserver, making data backups, does data verification and data cleaning. He has designed a number of applications for the HDSS including the data collection application for HDSS. He has also designed the query and validation application for data management as well as the entire software of Vadu HDSS. He speaks and writes English and Marathi.
- ***P8***. Participant 8 was a male who worked as the field coordinator (FC). He trains FRAs on the job, how to conduct interviews, enroll participants and prepare FRAs for field work. In summary he ensures the smooth running of field work. He speaks and writes English and Marathi.
- ***P9***. Participant 9 was a male who worked as the field supervisor (FRSs). He trains and monitors day-to-day activities of FRAs at the field site in the catchment area and ensures all outstanding queries are resolved while ensuring the smooth running of field work. He speaks and writes English and Marathi.
- ***P10***. Participant 10 was a male who works with the electronic system as a FRA in HDSS but had worked on other projects.
- ***P11***. Participant 11 was a male who works with the electronic system as a FRA in HDSS.
- ***P12***. Participant 12 was a female who works with the electronic system as a FRA in HDSS.
- ***P13.*** Participant 13 is a female who works with the electronic system as a FRA in HDSS. She worked on other projects on the hardcopy.
- ***P14***. Participant 14 is a male who works with the electronic system as FRA in HDSS.
- ***P15***. Participant 15 is a female who works with the electronic system as FRA in HDSS.
- ***P16***. Participant 16 is a female who had worked on a number of projects since 2009 on the manual system and currently working on HDSS project using the electronic system.
- ***P17***. Participant 17 is a female who works with the electronic system as FRA on HDSS project.

- ***P18***. Participant 18 is a male with over 6 years of experience as FRSs on HDSS and has experience on both the paper-based and electronic system but currently working on another project.
- ***P19***. Participant 19 is a male with 4 years of experience on both paper-based system and electronic system on HDSS project and had worked on other projects. He is currently a FRSs.
- ***P20***. Participant 20 is a female with over 10 years working experience and she had almost worked on every project. She worked as FRA and rose to FRSs in HDSS project and had worked on both paper-based and the electronic system.
- ***P21***. Participant 21 is a male who works as the HDM and responsible for the overall planning, designing and execution of all data management software applications and database management.

5.2. THEMES ARISING FROM DATA

The main themes that arose from data categorization and triangulation during the data analysis process[28-29] were the responses of participants in relation to lapses in the existing system at Vadu HDSS in data collection and data management among various roles and responsibilities in the DMT.

Theme 1: Current lapses in data collection at Vadu HDSS

Responses from participants with the same roles and responsibilities working at the field together with their supervisors and the coordinator reacted similarly. All the 13 FRAs, 4 FRSs and FC had the theme below:

- Problems from Respondents - They do not understand why certain information were required during enrolment and sometimes follow ups so they refuse to provide the right information. After watching the Vadu movie/documentary, they see the need to provide right information to FRAs/FRSs. They want some discounts from KEM hospital when they visit. Some respondents do not want ids to be written on their doors. Major problems confronted at the field are communication problems where some respondents find it difficult to understand certain

questions posed to them. They sometimes are unable to provide their right age or date of birth, or do not provide them at all.

- Problems from FRAs and FRSs - FRSs find it difficult locating FRAs when unannounced visits are made at the field. Sometimes it is a problem for some FRAs locating keys on keyboard. Some FRAs and FRSs lack experience using laptops. Some of the new FRAs find it difficult filling e-forms, these problems are mostly resolved after a week. Some FRAs complained that while filling the e-forms, if a mistake is committed the entered data cannot save. Some FRAs complained that all questions were mandatory on the electronic and tablet application which makes data entry difficult. Some FRAs complained that they had problems starting the laptops. Some FRAs complained that it was at times difficult shutting down the laptops. Some FRAs find problems accessing the HDSS data system. E-forms cannot be accessed due to technical problems. It was difficult searching names with minimum of 4 characters which were creating problems for names with 3 characters. Switching systems from electronic to manual creates functional problems which makes data collection process slow. Laptop batteries run down during data capture. FRAs sometimes need permission to update respondents' information.

Theme 2: Current lapses in data management at Vadu HDSS.

Participants with the different roles and responsibilities reacted differently. Below are the themes that arose:

P6 reported that there were no tools available for application on the tablet.

P7 stated that some respondents have problems spelling their names, and do not provide the right date of birth or age so matching data is difficult.

P21 stated that there was no detailed documentation of data management processes. Inadequate analysis was done on the datasets hence many unknown facts are hardly reported.

5.3. DISCUSSIONS

Although a number of lapses were outlined by the study participants among the various roles and responsibilities in data management life cycle, the HDM together with his team have gradually put in place measures to solve some of the problems identified through interview sessions and observation.

Data initially captured manually using the paper-based modalities usually are prone to data entry errors. In the manual system, informed consent form was not used in the enrolment process. Bulky materials were carried to the field like notebook, updating list for all villages and pens in a bag to the field daily. It was difficult and time consuming to search existing records for updating. More time was required to complete one questionnaire. Less validation checks were in place where FRAs manually entered date of visit, time of visit and their names. Physical monitoring of 10 to 12 FRAs were assigned to one FRSs, who visited the field un-announced to check if FRA was completing questionnaires correctly. All questionnaires were crossed-checked manually by FRSs to ensure that there were no blanks else those questionnaires with blank responses were queried and sent back for resolution. Physical checks were done on 10% of the filled questionnaires captured using the field quality control checks and back checks. Filled questionnaires were entered manually by data entry clerks or operators after which validation checks were run on the entered data and cleaned data is stored on sever for future referencing. The entered questionnaires were arranged by round numbers and village names and sent to an external archive for future referencing. This routine processes made data processing inefficient.

The DMT at Vadu HDSS introduced the use of laptops to collect data from the field. An application was developed to capture data electronically. This was to eliminate the work of the data entry clerks or operators and migrated from the manual system to the electronic system. Capturing data electronically was however a better option which had more advantages over the manual system or paper-based system. The merits of the electronic system using laptops are outlined below:

Questionnaires are manually filled if the laptop develops a fault at the field. 2 laptop batteries are sent to the field in case one battery runs down there would be a stand-by. Responses on the questionnaire are

captured using the laptop. Rate of committing errors are quiet minimal. Search for existing household head was quite easy, first 3 characters of first name and first 3 characters of last name of respondent is used in searching. Completing a questionnaire for a household with few members required less than 5 minutes, for households with many members required about 8 to 15 minutes to complete one questionnaire. Quick heal antivirus is installed on all the laptops and server which is automatically updated to protect the laptops and data from being destroyed provided there is internet access or connectivity. The EDC application on the laptop has a lot of validation checks as compared to the manual system. For instance, usernames and passwords are not required to start laptops. FRA logs unto the HDSS system using username and password. No FRA can log unto another laptop using their usernames and passwords. FRAs are encouraged to keep their passwords private. The system on the laptop generates date, time and FRA's name. The system generates the following features on all form types:

- Interview Start Time - Time in hours (06:00am to 21:00pm) and minutes, start time should be less than end time, time format should be 24 hours.
- Interview Date - Mandatory (Not Null), interview date should not be future date, interview date should be greater than date of birth.
- FRA name - Mandatory (Not Null), the system automatically displays FRA's name.
- Interview End Time - Mandatory (Not Null), time in hours (06:00am to 21:00pm) and minutes, end time should be greater than start time, time format should be 24 hours.

FRAs are trained to enter response directly unto the systems which is a merit over the manual system. FRAs can view Vadu HDSS entry screen and can edit entries made, but cannot delete any of the responses captured earlier. FRA notifies DM or FRSs for appropriate corrections when they commit certain mistakes like double responses for the same study participants. Validation checks (Field quality control checks and back checks) are in-built on the system and after data cleaning; data is saved and backed up unto the server. Weekly backup – Every Monday, all FRAs bring their laptops for backup and updates. Every laptop has automatic backup and saved on the local drive D. Data export from FRAs'

laptop to DM's laptop is saved unto the Dump Database. Validation checks are run on the dumped database and portable database on the server is updated. Data from Portable Database is exported from the DM's laptop to FRAs 'laptop. Portable Database does not have up-to-date data but lacks a week's data. After cleaning the data on the local drives, weekly updates are saved unto local server and Dropbox.

With all the measures put in place, there were still some problems which the laptops could not solve in the current system. A HDSS software application was therefore developed with more inbuilt validation checks on tablets. This was a proposed system as at the time the research work was carried out. There are however some advantages of using the new system. Fewer materials would be carried to the field, a bag containing a tablet with 5 different types of all questionnaires and pens. The GPS android application on the tablet is activated to track and capture the location, date and time during field work unknown to both FRAs and study participants. An inbuilt audio recording is also available on the tablet application to record all interviews between FRAs and respondents. Rate of committing errors is minimal because the FRAs must selects appropriate values from the options provided in the system which help to reduces errors. From the tablet application, respondents or participants could append their signatures by placing their finger or thumb prints on screen of the android application on tablets. All questions on the e-form of the application are mandatory so if there are blanks, the responses captured will not be saved if participants do not provide answers to all questions, data captured cannot be stored on the HDSS application. The tablets proposed to capture data are brand new hence reduces the tendency of hardware problems. Audit trails have been incorporated into the data collection and management routine process to capture the number of errors, time required for responses of study participants to be captured. FRAs can upload data captured on the HDSS system at the close of work daily when responses of a household are complete to the server without necessarily physically present at the office if there is the availability of network. Training sessions are conducted for FRAs, FRSs and FC before the commencement of a

round. The android application on the tablet would make data management routine process much efficient if fully implemented.

CHAPTER SIX: INFORMATION ASSURANCE AND DATA QUALITY SOPs FOR VADU HDSS SITE

This chapter outlines the proposed information assurance policy (IAP) found in Appendix Four and the Standard Operating Procedures (SOPs) found in Appendix Five for Vadu HDSS. It is intended that steps in the SOPs document is replicated in data management life cycle activities at HDSS sites for research data quality assurance.

6.1 VADU HDSS DATA LIFE CYCLE

Figure 6.1 below shows the data life cycle at Vadu HDSS. It explains how research data is captured, processed and analyzed for every cycle of the DSS data capture round. All preliminary checks on laptops and tablets for data collection are done before the commencement of field work. Data on the laptops and tablets are synchronized with the main data repository. FRAs conduct interviews at households within the surveillance area and capture data from study participants using e-forms available on the laptops and tablets. FRSs make unannounced visits on households to conduct random quality checks on data captured by FRAs. If all quality checks conducted are passed, the data is dumped into the secondary database. This is followed by automated quality checks performed against the rules. If all quality checks conducted are passed, data is stored in a primary HDSS database. Data versions are then assigned and sent to archives for safe keeping and future references. If the data quality checks conducted fails, then data is returned to FRAs through FRSs for necessary corrections. DSS datasets goes through the same process until it passes the test. Data can only be released for analysis based on formal request and the approval of the Head of Vadu HDSS through the HDM and/or DM.

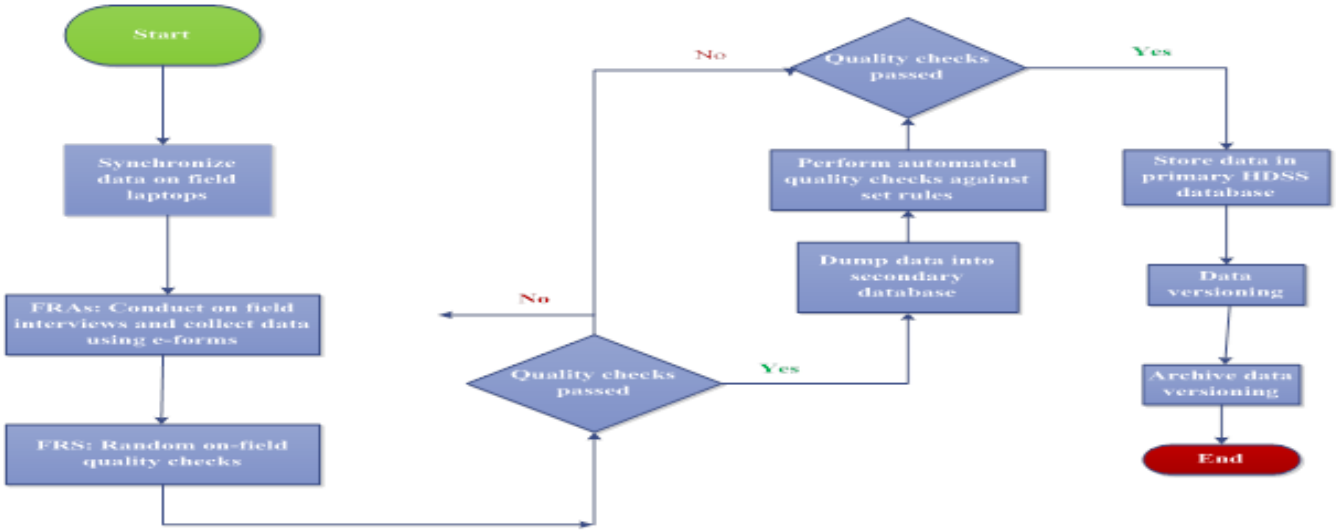


Figure 6.1. Research Data Life Cycle at VADU HDSS

6.2 INFORMATION ASSURANCE POLICY

The proposed IAP document for Vadu HDSS is to guide data management life cycle activities which could be found in Appendix Four. This document is made of a number of policies for all members of the DMT to ensure the quality of data during research activities at Vadu HDSS. The IAP document is divided into 2 main sections which are the general policies and the data management policies. The general policies outline all policies DMT members are encouraged to comply. The sections of the general policies are as follows:

Wearing of badges is to make it easier to identify members of DMT from other staff. Authentication policy encourages members to secure data and all computers, laptops, tablets and servers by logging unto them with their usernames and passwords. Password creation policy ensures that members should be assigned unique passwords to access system-level privileges based on their roles and responsibilities. Password changing policy recommends that users change their passwords at least every six months. Password protection policy provides certain guidelines for members to create their passwords, for example members should not write down their passwords for security reasons. Prohibited activities by DMT policy provides a list of activities that should not be done by members.

This policy is to protect research data from unauthorized access like phisher attack and/or threat from an insider, prevent data breaches and provide a form of security for the data collection and data processing devices such as workstation, laptops and tablets[40].

User privileges policy encourages that when a DMT member's roles and responsibilities changes the access system-level privileges also changes. Termination of user logon account policy outlines that when a member's contract ends, their passwords are made inactive and all company's belongings must be returned. This is to make sure that DMT member who leave Vadu HDSS permanently do not get access to any dataset and other stuff belonging to the research centre; their passwords are deactivated[41].

Data security policy steps are guidelines which enables data to be protected from unauthorized persons. Hard**w**are and software security protections policy recommends the use of antiviruses, security locks and screen lock to secure all software and hardware. Backup policy recommends that automatic backups have been made available on the laptops, computers and tablets unknown to users which have been encrypted. Hardcopy reports and paper work policy encourages members never to leave paper records around their work area and lock all paper records in a file cabinet at night or when you leave your work area. Also all papers with sensitive information be shredded before being disposed. Wireless usage standards and policy encourages members to consult ITO or HDM for appropriate access accounts. Software requirements policy prevents members to install any software packages except what have already been installed on the device assigned to them based on their roles and responsibilities. Any member found to have violated this policy may be subject to disciplinary action. It is very crucial to have policies to guide wireless usage due to the security related problems with wireless networks, they transmit potentially sensitive information over the airwaves which can pose a huge threat to Vadu HDSS network. This means that the information flowing across the network can be intercepted by anyone within range who has a laptop equipped with a wireless network card and is easier for hackers to enter the network without having to deal with the constraints normally associated with an Internet based attack[41]. If no policy or standards are developed to guide wireless usage at Vadu HDSS, unauthorized

staff and even visitors are likely to access data easily and make deliberate or unconscious changes to data and probably. Staff and even visitors are probable to visit certain virus prone websites which may introduce certain viruses unto the Vadu system. Staff should contact the HDM, ITO and/or DM to install any additional software, this could prevent the system from been slow, and prevent required software packages for processing data conflict with unauthorized packages[42].

The next section outlines data management's policy which is as follows:

Informed consent policy which encourages all members designated at the field to obtain informed consent from all study participants before enrolment. Audit trails and controls policy to guide members to ensure that DMP and study protocol are followed to produce quality data through monitoring and audit trail recordings. Training policy recommends training for all members designated at the field before a new round or project begins. Field monitoring policy ensures certain mechanisms to be adopted by FRSs, DM, and/or SADs to design and monitor field work by capturing locations and correct responses from study participants. QC/QA policy recommends that all QC and QA steps are implemented in all data management life cycle in accordance with DMP and study protocol. Query resolution[43] policy ensures that all queries are resolved in accordance with DMP and study protocol to reduce errors.

Informed consent policy encourages free participation, informed consent should be obtained from every participant prior to study participation. All study data should be recorded, processed and stored in a way that allows its accurate reporting, interpretation and verification. The confidentiality of participant records should be protected, respecting the privacy and confidentiality rules of the applicable regulatory authority[44]. The essence of audit trails and controls policy to keep track and logs or documents to determine whether these activities were conducted. This is to also determine whether the data were recorded, analyzed, and accurately reported according to the study protocol, standard operating procedures, good clinical practice and the applicable regulatory requirements [26]. Is it very important

to training all DMT members adequately to capture required data according to the study protocol, this will reduce potential errors at data collection and management and help to reduce load on data cleaning QC/QA policy ensures that routine techniques and activities undertaken within the quality assurance system to verify that the requirements of the study-related activities have been done. This involves the proof of the availability of signed consent forms, study protocol is followed adequately, data is handled accordingly, cross check source data against database, ensure completeness and correctness during data entry process, prevent duplicates, regular validation checks; proof of an appropriate audit trail mechanisms, regular automatic and manual data backup mechanisms. Adequate monitoring mechanism should be adopted to produce clean and quality data ready to be analyze[45].

Query resolution policy is to ensure that all queries are resolved appropriately[43].

Validation checks and cleaning policy ensures that HDM, SAD and/or DM design a validation check list to help reduce errors, ensure accuracy and completeness of data on all e-form types on the HDSS application in accordance to DMP and study protocol. Data storage policy ensures that processed data is kept safely for future retrieval at the end of every round in case of any system failure after data cleaning[46-47]. This policy helps data to be cleaned and prepared for data analysis and data is archived and kept for future referencing.

Data access and sharing policy promotes data sharing and enable reliable and registered access to data for research stored electronically, educational purposes and publications[47]. It is expected that before publication, the data are accessed under the data sharing and access guidelines. This policy also ensures that data is anonymized to hide study participants' identifiers. Datasets are labelled with standard identifiers and versions so that users can easily differentiate and compare separate analyses of datasets. Researchers should be able to link associated datasets stored in various databases and link datasets to any publications based on the data[63][62][54]. Details to this policy can be found in Appendix Four.

From the theme that arose (Current lapses in data management at Vadu HDSS), the HDM stated that there was no detailed documentation of data management processes. Inadequate analysis was done

on the datasets hence many unknown facts were hardly reported. He recommended the need for process documents like an SOP which is an important document for streamlining processes to have quality data collection and data processing.

The IAP and SOP will provide guidelines for routine data management life cycle activities in the DMT since the centre do not have this in place. This will help DMT member produce quality and reliable research data starting from data collection at the field through data processing where datasets are prepared into analytical datasets.

6.3 STANDARD OPERATING PROCEDURE (SOP) PROPOSED FOR VADU HDSS

The proposed SOP document for Vadu HDSS is made up of steps, guideline and instructions of how to perform the policies in the IAP to guide data management life cycle activities could be found in Appendix Four. The steps in this document are to be observed by all members of the DMT members to ensure the quality of data in research activities at Vadu HDSS. The sections in the IAP are same in the SOP document but goes further to give steps, guidelines and/or instructions to perform those policies. Below are the parts of the SOP document summarized and details could be found in Appendix Five:

Section 1: **Authentication - Password Creation, Change and Protection**

This outlines standard to guide DMT members at Vadu HDSS in creating strong passwords, protecting those passwords and the frequency of change. This is to protect research data and company's assets. DMT member require a strong password to gain access to the Vadu system. The following are guidelines to be considered when creating a strong password:

Password should contain at least 12 alphanumeric characters, contain both upper and lower case letters, contain at least one number, contain at least one special character, password should not be found in a dictionary, should not have personal information, should not contain number patterns, should not have words spelled backward, or preceded or followed by a number[29] [48-49] (See Appendix Five).

Section 2: **Informed Consent**

This outlines the processes for obtaining written informed consent for enrolling participant into the Vadu HDSS study, this is a requirements applicable to all DMT members designated at the field to obtain informed consent from all study participants prior to the commencement of the study. The following are the guidelines to be considered to obtain informed consent from study participants:

Permission must be granted by household head before obtaining informed consent from potential study participants. FRAs, FRSs and FC must give details of the study, purpose of the study, expected duration of participation and the procedures that will be followed. Risk and Benefits disclosure should be made known to potential study participants. Any available courses of treatment that might be beneficial to the participants must be disclosed prior to the study. Potential study participants must be assured that their records shall be kept private and confidential. Any compensation made available to potential study participants must be disclosed prior to the study. Study participants should understand that partaking is voluntary, refusal to participate or withdrawal in the course of the study will not involve any penalty or loss of benefits. Potential study participants must be assured that the results of the study would be related to them when available. A written informed consent in the dialect well understood by study participants will enable them make informed decisions whether or not to participate in the study. Informed consent form should be filled and signed or thumb printed by study participants. Participants signature or thumb print can be captured electronically on the HDSS application available on the tablets[52].

Section 3: Audit Trail

This is to certify that all DMT members at Vadu HDSS should follow while performing their routine activities to produce quality research data through monitoring and audit trail recordings in accordance to DMP. The following are some of the instructions:

SAD and DM will incorporate some mechanisms of system online monitoring and audit trail recording, protecting, reviewing and reporting security breaches or anomalies to the HDM.

SAD and or DM should occasionally monitor online programmer activity to ensure audit trail functions are operating and reports are reviewed weekly and it should be able to aid SAD and/or DM to reconcile audit trail anomalies. ITO/DM shall enable event auditing on all computers that process, transmit, and/or store research data for the purposes of generating audit logs. Audit trails shall be stored on separate workstations to reduce the impact of audits trails or logs being accessed on individual laptops and tablets. HDSS audit files shall be stored in a locked room and kept according to protocol. HDM/ITO/DM/SAD shall be responsible for monitoring audit trails to check if there are protocol violations and any inappropriate validations[51].

Section 4: Training

To improve and/or produce quality research data, DMT at the field (FRAs, FRSs and FC) collecting data must be trained to use the HDSS application appropriately on the laptops and/or tablets to capture all required responses from study participants during enrolment of participants and various event updates. The HDM and/or DM must ensure all members understand the entire data collection process, verify the accuracy and completeness of data in compliance to study protocol and DMP. The following are the steps for training:

HDM and/or DM should organize training sessions for all DMT (FC, FRSs and FRA) on the field to understand how the HDSS application works. FC and FRSs must ensure that all FRAs have made themselves available for the training sessions at the beginning of each round and subsequent training sessions. Each member of the DMT responsible for collecting data must be fully trained on the HDSS application on both laptops and tablets prior to the beginning of the study. Each member of DMT will have proper security privileges assigned prior to entering data into the HDSS application. No member of DMT with security privileges will grant access to another person under their identity and password. Only members of DMT with proper security access will have access to the HDSS application. Different privileges shall be granted to FC, FRSs and FRAs. A paper or oral based assessment should be conducted to ascertain the level of understanding of the training undertaken[52].

Section 5: Field Monitoring

To ensure that FRAs are capturing data correctly in accordance with Vadu HDSS DMP and study protocol, all members of the DM ensures that all FRAs are tracked while conducting their routine activities on the field. DMT members at Vadu HDSS have in place the following guidelines:

- FRAs must design a weekly work plan and submit to FRSs and/or FC. With the weekly work plan, FRSs and/or FC knows the location where each FRA is. 12 FRAs are assigned to one FRSs. FRSs must visit about 60% of FRAs in a week.
- Phone calls - FRSs and/or FC calls the FRAs in case they are not found in the expected location for direction because some households are far apart.
- FRSs and/or FC randomly cross-checks responses captured on the HDSS application with study participants to ensure whether data captured is accurate.
- HDSS applications on the tablets have been programmed to make a voice recording of the interactions between FRAs and study participants during data collection but the voice recording is unknown to FRAs.
- GPS application on tablets have been activated and programmed to capture location where the interviews and data collection was done but unknown to FRAs.
- FC visit field to ensure smooth running of all field work with accordance to DMP and study protocol.

Section 6: Quality Control(QC)/Quality Assurance (QA)

This is to ensure that the study is performed and the data generated, documented and reported in compliance with GCP and the applicable regulatory requirements. All members of the DMT should adhere to this document to produce quality data. Below are the steps:

- HDM and/or DM must ensure that all required QC/QA processes are followed in accordance with HDSS study protocol and validation checklist of DMP. This involves checking protocol with electronic data by ensuring that:

- All validation checks are run on all captured data. If more errors are detected after validation checks by some FRAs who captured data, training sessions are organized for them by DM.
    - The HDSS application on the laptops and tablets has been programmed to capture all required responses else the e-form cannot be saved.
    - The HDSS application would not be saved if it detects protocol violation, any missing values, outliers (range checks) and inconsistencies [58].

Section 7: Query Resolution Guidelines

This is to promote consistent, efficient and effective data management life cycle. DMT must be trained to resolve queries in accordance with Checks daily for queries. All queries must be resolved within 2 weeks unless specified by study protocol. If more information is required to resolve queries, the study protocol should be referred. Study participants will have to be consulted for queries with exceptional problems. DM will make reference to unanswered query list with FC/FRSs for outstanding queries weekly [58].

Section 8.0: Validation Checks and Cleaning

HDM, SAD and/or DM at Vadu HDSS have designed a validation check list to help reduce errors, ensure accuracy and completeness of data on all e-form types on the HDSS application in accordance to DMP and study protocol. This document applies to all members of the DMT at Vadu HDSS. The following are the guidelines:

Every DMT member must log unto the HDSS systems using assigned username and password.

Every laptop and/or tablet is assigned to only one field worker. FRAs are encouraged to keep their passwords private and secret. System generates date and time – The HDSS application generates date and time automatically like the following: Interview Start Time, Interview Date, Field worker name and Interview End Time, other guidelines are as follows:

Ensuring accuracy and completeness, Account Privileges: Limited Access - FRA can only view, edit and update entry screen on HDSS application but cannot delete responses entered earlier. Full Access

– HDM/DM have full access to the system. They can view, edit, update and delete. Query Resolution - The HDSS application prompts DM for outstanding queries which must be resolved. Field QC and Back Checks - FRSs runs the field quality control checks and back checks done on data captured. Final Checks - After query resolutions, validation checks are executed finally to ensure data is cleaned. Audit Checks - If DM detects that responses captured have many errors, or responses are captured less than an average time, then DM suspects FRAs for improper evaluations and FRAs are questioned for explanation.

HDM and/or DM must read and understand study protocol to ensure protocol is in conformity with DMP and validation checklist. This is required for appropriate updates to the HDSS application[52].

Section 9: Data Storage and Archiving

These are measures put in place to protect data at Vadu HDSS. These measures are backup and storage mechanisms. These are put in place to secure data for future retrieval at the end of every round in case of any system failure after data cleaning. All members of the DMT should adhere to this SOP to secure data for easy retrieval. FRAs are required to upload the responses captured on the HDSS system at the close of work daily to the server without necessarily physically present at the office if there is the availability of network. The clean data on the local and cloud server is updated weekly. Copies of clean data are uploaded unto every individual laptop and tablet for the next week field visits.

Reports are generated weekly by the FC aided by DM[53].

Section 10: Data Access and Sharing

This outlines the processes DMT at Vadu HDSS have in place for data to be accessed and shared with the public. The HDM and DM at Vadu HDSS should adhere to these processes for data access and sharing with the public[54].

6.4 RECOMMENDATIONS

ISACA/COBIT 5 family products and ISO/IEC ISMS as well as other standards were used as benchmark to design IAP and SOPs for the Vadu HDSS site. The IAP and SOP have not been tested on pilot bases.

These processes could be simulated in an ongoing research to ascertain its impact. SOP may be subject to updates with suitable versions and date to suit other data management processes due to changing technologies. This is to ensure high data quality, validity and integrity of data, and provide adequate data protection. Training sections and the distribution of the proposed IAP and SOP documents would be post-research activities.

CHAPTER SEVEN: CONCLUSION AND FUTURE DIRECTIONS

This chapter presents summary of the study, appraises the limitations of the study and finally conclusion.

7.1 CONCLUSION

The findings from this study from different roles and responsibilities in data management life cycle revealed that a lot of measures have been put in place to improve data management processes and also reduce existing lapses at Vadu HDSS. The study participants freely shared their experiences with the investigator and the response rate was 95.5%. Interview guide sought to identify lapses with data management processes and information quality assurance at Vadu HDSS. There were 6 sections in the interview guide ranging from section A to F which are outlined below: section A (Roles and Responsibilities), section B (Data collection, data processing, data preparation for analysis and audit trail), section C (Validation mechanisms), section D (Data storage, filing and archiving), section E (Audit Assurance), section F (Information Assurance).

Responses from roles and responsibilities revealed these roles and responsibilities: FRAs, FC, FRSs, ITO, SADs, DM and HDM. The data entry supervisors' roles and responsibilities were listed on the interview guide but not found in the current system, rather, there were more roles assigned to members of the DMT. The new roles identified were FRAs, ITO and SADs. Personnel at different data management stages may differ according to sites but this may not affect research data quality.

Responses from data collection, data processing, data preparation for analysis and audit trail showed that the data collection processes were done electronically and also when the electronic device developed a fault, they switch to the manual system. Participants responded positively that there were some measures in place to ensure that data collected is accurate. These measures include physical monitoring by the FRSs using field QC and back checks, the electronic monitoring system found on the in-built validation checks, GPS activation and automatic audio recordings found on the android device. Follow-

ups were made due to the previous information on visit. This finding implies that quality control checks have been put at all levels to ensure quality research data.

Furthermore, on validation mechanisms, logging unto the HDSS application with assigned usernames and passwords were some of the validation mechanisms in the current system. This is an important finding since the use of usernames and passwords considerably ensures data safety.

As regards data storage, filing and archiving the source documents (manual/paper-based system) were sent to external archiving for future referencing and processed data (cleaned data) are kept on the local server and Dropbox, as well as other storage devices for future retrieval or referencing. Vadu HDSS have in place effective mechanisms for data storage and for future data references.

Also on audit assurance, inappropriate evaluation could be detected by logging unto the system with usernames and passwords. Audio recordings are captured during data collection and activated GPS on the tablet application, when database is manually opened, in-appropriate evaluation could be detected. Field QC and back checks, re-checking information captured by FRAs with respondent for accuracy and some undocumented mechanisms are in place to compare results with other researchers and published papers and abnormalities are reported. This is an indication that effective audit assurance mechanisms have been put in place in data collection and management processes.

According to responses from information assurance, users logged unto the system using usernames and their encrypted passwords to gain access to HDSS application on the electronic device. This ensure authentication in the data management processes. After data cleaning and updates, data is backed up on both the local server and Dropbox. Without user authentication it would be difficult to get a copy of another backup copy from the automatic backup copy. Processed data stored on servers is accessible to only authorized personnel. Responses revealed that without correct answers to questions the system is unable to save data entered, so all questions were mandatory (No blanks) and incomplete data was not possible to be captured on the android application. FRAs must choose from options provided on application to reduce manual data entries to lessen errors. In-built automatic validation checks were

available on HDSS application. Response showed these measures were put in place to ensure data correctness and completeness in the data management process in the current system. Users logged unto the HDSS application in the current system and that is how data privacy and confidentiality issues were handled. Currently, Vadu HDSS have no written policy and SOP to guide data management life cycle activities. We adopted the ISACA/COBIT 5 guidelines and ISO/IEC ISMS as benchmark to develop an IAP and SOPs to guide data management life cycle activities with the aim of producing/augmenting quality research data and steps in SOP could be followed to reproduce similar outcomes [10][12][13][64].

REFERENCE

1. Sankoh O, Sharrow D, Herbst K, Kabudula CW, Alam N, Kant S, et al. The INDEPTH standard population for low- and middle-income countries, 2013. Glob Health Action. 2014; 7:23286. Available at: http://dx.doi.org/10.3402/gha.v7.23286
2. Ingole V, Juvekar S, Muralidharan V, Sambhudas S, Rocklöv J. The short-term association of temperature and rainfall with mortality in vadu health and demographic surveillance system: A population level time series analysis. Glob Health Action. 2012;5:44–52.
3. Krishnankutty B, Bellary S, Kumar NBRR, Moodahadu LS. Data management in clinical research: An overview. Indian J Pharmacol [Internet]. 2012 Mar [cited 2014 Sep 20];44(2):168–72.Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326906&tool=pmcentrez&rendertype=abstract
4. Gerritsen MG, Sartorius OE, vd Veen FM, Meester GT. Data management in multi-center clinical trials and the role of a nation-wide computer network. A 5 year evaluation. Proc Annu Symp Comput Appl Med Care. 1993;44:659–62. Available at: http://www.ijp-online.com/crossrefCitation.asp?doi=10.4103/0253-7613.93842
5. Nugroho H. Conceptual model of IT governance for higher education based on COBIT 5 framework. J Theor Appl Inf Technol. 2014;60(2):216–21.
6. Ngouongo SMN, Löbe M, Stausberg J. The ISO/IEC 11179 norm for metadata registries: Does it cover healthcare standards in empirical research? J Biomed Inform. 2013;46:318–27.
7. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Summits Transl Sci Proc. 2010;2010:1–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21347133
8. Sachdeva S, Bhalla S. Semantic interoperability in standardized electronic health record databases. Vol. 3, Journal of Data and Information Quality. 2012. p. 1–37. Available at: http://dl.acm.org/citation.cfm?doid=2166788.2166789
9. du Preez G, Hardy G, Lainhart JWI. Implementing and continously improve IT governance. Inf Secur. 2009;1–74. Available at: http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Implementing-and-Continually-Improving-IT-Governance1.aspx/%209781604201192
10. Hardy G. Using IT governance and COBIT to deliver value with IT and respond to legal, regulatory and compliance challenges. Inf Secur Tech Rep Leg Regul Compliance Asp Inf Secur. 2006;11(1):55–61.
11. Haux R, Knaup P, Leiner F. On educating about medical data management - the other side of the electronic health record. Methods Inf Med [Internet]. 2007;46:74–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17224986
12. Meester GT, Sartorius OEH, Fanggidaej D, Gerritsen MGJM. CADANS-a customized network for cardiology in the Netherlands. Images Twenty-First Century Proc Annu Int Eng Med Biol Soc. 1989; Available at: https://www.researchgate.net/publication/3552055_CADANS a_customized_network_for_cardiology_in_the_Netherlands/doi: 10.1109/IEMBS.1989.95667
13. Sajdak R, Trembath L, Thomas KS. The importance of standard operating procedures in clinical trials. J Nucl Med Technol [Internet]. 2013 Sep 1 [cited 2014 Sep 14];41(3):231–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23853088

14. Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, et al. Heterogeneity prevails : the state of clinical trial data management in Europe - results of a survey of ECRIN centres. 2010;1–11.

15. Young B. CS361C : Information Assurance and Security. 2015;1–62. Available at: http://www.cs.utexas.edu/~byoung/cs361c/slides1-intro.pdf

16. Van Grembergen W, De Haes S. COBIT as a Framework for IT Assurance. In: Enterprise Governance of Information Technology [Internet]. 2009. p. 165–82. Available at: http://www.springerlink.com/index/10.1007/978-0-387-84882-2

17. De Haes S, Van Grembergen W, Debreceny RS. COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities. J Inf Syst [Internet]. 2013;27:307–24. Available at: http://dml.regis.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=88181972&site=ehost-live&scope=site

18. Ximenes P, Santos A, Jr JC. SecMD (Secure Medical Database)*. 2003; Available at: http://academic.research.microsoft.com/Publication/61464743/improved-security-of-audit-trail-logs-in-multi-tenant-cloud-using-abe- schemes/doi:10.14569/IJACSA.2014.051120

19. Annas GJ. HIPAA regulations - a new era of medical-record privacy? N Engl J Med. 2003;348(15):1486–90.

20. R GM, Cash R. Research involving medical records review : an Indian perspective. 2015;III(2):55–7. Available at: http://www.issuesinmedicalethics.org/index.php/ijme/article/view/650/1624

21. National Institutes Of Health. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule. HIPAA Privacy Rule Information for Researchers.2003;2800228032: 1-32. Available at: https://privacyruleandresearch.nih.gov/pdf/hipaa_privacy_rule_booklet.pdf

22. Guerra SÁ. Harmonization of quality standards for clinical trials. ISO-9001 standard and guide of good clinical practice . Armon estándares Calid para ensayos clínicos norma ISO 9001-guía buena práctica clínica [Internet]. 2011;45:380–8. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-80054123296&partnerID=40&md5=b8173169c215a2a840730da1868cf7e9

23. Burnstead B, Furlan G. Unifying drug safety and clinical databases. Curr Drug Saf [Internet]. 2013;8:56–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23656448

24. Sheikhpour R, Modiri N. An approach to map COBIT processes to ISO/IEC 27001 information security management controls. Int J Secur its Appl. 2012;6:13–28.

25. Jiang K, Cao X. Design and implementation of an audit trail in compliance with US regulations. Vol. 8, Clinical Trials. 2011. p. 624–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21900341/doi: 10.1177/1740774511413943

26. National Institutes of Health. The Belmont Report. Belmont Rep Ethical Princ Guidel Prot Hum Subj Res [Internet]. 1979;4–6. Available at: http://www.salesianos-cadiz.com/Descargas/Escolar/ESO-Bachillerato/Bachillerato/2do Bachillerato/2do A/InformeBelmont.pdf

27. Health U, Of E, Of N. A Step-by-Step Guide to Qualitative Data Analysis. Vol. 1 Issue 1, p63. Available at: http://connection.ebscohost.com/c/articles/14860050/step-by-step-guide-qualitative-data-analysis

28. Bricki N, Green J. A Guide to Using Qualitative Research Methodology. Med Sans Front [Internet]. 2007;11–3. Available at: http://msf.openrepository.com/msf/handle/10144/84230

29. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Research design Qualitative quantitative and mixed methods approaches. 2014. 398 p.doi:10.1007/s13398-014-0173-7.2 Available at: http://isites.harvard.edu/fs/docs/icb.topic1334586.files/2003_Creswell_A%20Framework%20for%20Design.pdf

30. ISO / IEC 27001 Information Security Management System. 2008; Available at: http://www.iso.org/iso/home/standards/management-standards/iso27001.htm

31. Iso Iec. BS ISO/IEC 27002:2005 Information technology — Security techniques — Code of practice for information security management [Internet]. ISO. 2005. p. 130. Available at:http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50297

32. ISACA. IT Assurance Guide. Management. 2009. 269 p./isbn: 1933284749. Available at:http://www.inf.unideb.hu/~fazekasg/oktatas/cobit/IT%20Assurance%20Guide%20-%20CobiT%204.1%20(2007).pdf

33. Lainhart J, Oliver D. Integrating ISACA Frameworks Into One Overarching Framework: COBIT 5. COBIT Focus [Internet]. 2010;2010:6–9. Available at: http://ezproxy.library.capella.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=49765531&site=ehost-live&scope=site

34. Headquarters I, Meadows R. IS Audit and Assurance Guideline 2208 Sampling IS Audit and Assurance Guideline 2208 Sampling 1 . Guideline Purpose and Linkage to *Standards. ITAF™: A Professional Practices Framework for IS Audit/Assurance, 3rd Edition* 2014;2014:98. Available at: http://www.isaca.org/Knowledge-Center/Research/Documents/ITAF-3rd-Edition_fmk_Ukr_1214.pdff

35. Ruiz EDD, Marcos SO. Comparativa de ITIL v3 con COBIT 4 . 1 y desarrollo de una aplicación ITIL para el iPhone. Univ Pontif COMILLAS Esc TÉCNICA Super Ing Ing EN Organ Ind Proy. 2011;1.

36. Vadu Rural Health Program, K.E.M. Hospital Research Centre, Pune. Available at: http://www.kemhrcvadu.org/index.php/about-us/2014-03-02-11-12-33/hdss

37. Vadu Rural Health Program, K.E.M. Hospital Research Centre, Pune. Available at: http://www.kemhrcvadu.org/index.php/about-us/history

38. Randolph C. Barrows Jr. MD, Paul D. Clayton PhD. Privacy , Confidentiality : and Electronic Medical Records Abstract The enchanced Goals of Informantional Security In Health Care. J Am Med Inform Assoc. 1996 Mar-Apr; 3(2):139-48. Available at: http://jamia.oxfordjournals.org/content/3/2/139

39. Harrington JL. User and Password Security. In: Network Security: A Practical Approach [Internet]. 2005. p. 205–24. Available at: http://www.sciencedirect.com/science/article/pii/B9780123116338500092

40. DeLeon, P. (1983), POLICY EVALUATION AND PROGRAM TERMINATION. Review of Policy Research, 2: 631–647. doi: 10.1111/j.1541-1338.1983.tb00793.x. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1541-1338.1983.tb00793.x/abstract

41. Lehr W, McKnight LW. Wireless Internet access: 3G vs. WiFi? Telecomm Policy. 2003;27(5-6):351–70.

42. Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. In: Methods of Information in Medicine. 2009. p. 408–13.

43. Bernat JL. Informed consent. Muscle Nerve. 2001;24:614–21.

44. Geboy BNJ, Engle MA, Survey USG. Quality Assurance and Quality Control of Geochemical Data : A Primer for the Research Scientist. Vols. 2011–1187, USGS Open File Report. 2011.

45. Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Métro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies - The Pre-Optimon study. Contemp Clin Trials. 2011;32:16–24.

46. Walport M, Brest P. Sharing research data to improve public health. Lancet. 2011;377(9765):537–9.

47. Segalstad SH. Quality assurance of computer systems. What is needed to comply with ISO 9000, GMP, GLP, and GCP? Lab Autom Inf Manag. 1995;31:11–24.

48. McCollum T. ISACA UPDATES COBIT FRAMEWORK. Intern Audit [Internet]. 2012;69:15. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=79662718&site=ehost-live

49. Version D. KEMHRC Pune Data Access and Sharing Policy. 2014;

50. WMA General Assembly. World Medical Association Declaration of Helsinki. World Medical Association. 1964.

51. Broderick JS. ISMS, security standards and security regulations. Inf Secur Tech Rep. 2006;11(1):26–31.

52. Grandison T, Bhatti R. HIPAA compliance and patient privacy protection. Stud Heal Technol Inf. 2010;160(Pt 2):884–8.

53. Cutcliffe JR, McKenna HP. Expert qualitative researchers and the use of audit trails. Vol. 45, Journal of Advanced Nursing. 2004. p. 126–33.

54. Idänpään-Heikkilä JE. WHO guidelines for good clinical practice (GCP) for trials on pharmaceutical products: responsibilities of the investigator. Ann Med. 1994;26:89–94.

55. Ohmann C, Kuchinke W, Canham S, Lauritsen J, Salas N, Schade-Brittinger C, et al. Standard requirements for GCP-compliant data management in multinational clinical trials. Trials. 2011;12:85.

56. Rausher MD, McPeek MA, Moore AJ, Rieseberg L, Whitlock MC. Data archiving. Evolution. 2010;64(3):603–4.

57. Centers for Disease C, Prevention. HIPAA privacy rule and public health. Guidance from CDC and the U.S. Department of Health and Human Services. MMWR - Morb Mortal Wkly Rep. 2003;52 Suppl:1–17.

58. Kaur S, Choy CY. Ethical Considerations in Clinical Trials: A Critique of the ICH-GCP Guideline. Dev World Bioeth. 2014;8731(October 1990):23–4.

59. ISACA. Chapter VI: Audit/Assurance Program BCM Policy, standard and procedures. ISACA. 2011;25.

60. Yellowlees PW, Harry D. Standards for data collection and monitoring in a telemedicine research network. J Telemed Telecare. 2006;12 Suppl 2:S72–6.

61. Arredondo P, Ivey A, Sue DW, Parham T, Sue DW, Mio JS, et al. Guidelines on multicultural

education, training, research, practice, and organizational change for Psychologists. Am Psychol [Internet]. 2003;58(5):377–402. Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.58.5.377

62. Standards A. IT Standards , Guidelines , and Tools and Techniques for Audit and Assurance and Control Professionals Current as of 16 August 2010. Prof Ethics. 2010;(August):329.

63. Disterer G. ISO/IEC 27000, 27001 and 27002 for Information Security Management. J Inf Secur [Internet]. 2013;4:92–100. Available at: 10.4236/jis.2013.42011\nhttp://search.ebscohost.com/login.aspx?direct=true&db=i3h&AN=89254050&site=ehost-live

64. Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, et al. Standard operating procedures for serum and plasma collection: Early detection research network consensus statement standard operating procedure integration working group. J Proteome Res. 2009;8(1):113–7.

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 MIEKS F.N. TWUMASI (Student number: 709731) am a student registered for the degree of MSC IN EPIDEMIOLOGY 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: March, 29th 2016

APPENDIX TWO: QUESTIONNAIRE/INTERVIEW GUIDE

This seeks to identify gaps with data management processes and information quality assurance.

This could take about 40 minutes.

Section A (**Roles and Responsibilities**)

1.0 What are the roles and responsibilities in data management life cycles?

1.1 What is your role in the data management life cycle? Please tick one □ Head of Department □ Field Supervisor □ Data Entry Supervisor □ Data Manager □ Field Coordinator

1.2 Describe the flow of data management task.

Section B (**Data collection, data processing, data preparation for analysis and audit trail**)

2.0 How is data captured from participants on the field? □ Manual □ Electronic

2.1 Do you have any measures in place to ensure data collected is accurate? □ Yes □ No

2.1.1 If yes to question 2.1, please describe the measures?

2.2 What monitoring mechanism(s) do you have in place to monitor and document data management life cycle processes at various stages?

2.3 Are there any mechanisms in place to guide follow-up activities from the field? □ Yes □ No

2.3.1. If yes to question 2.3, please describe the process.

Section C (**Validation mechanisms**)

3.0 Are there any validation mechanisms before any data management activity? □ Yes □ No

3.1 If yes to 3.0, please describe the validation mechanism.

Section D (**Data storage, filing and archiving**)

4.0 How are source documents and processed data kept for retrieval or referencing?

4.1 What lapses you know of in data management processes that can affect data quality?

Section E (**Audit Assurance**)

5.0 Are there any mechanisms in the current system to detect personnel performing inappropriate evaluation? □ Yes □ No

5.1 If yes to question 5.0, please describe the process.

5.2 Are there any mechanisms in the current system to identify personnel whose objectivity is compromised? □ Yes □ No

5.2.1 If yes to question 5.2, please describe the process.

5.3 Are there any mechanism in the current system to detect the potential threat that affect the will of personnel as a result of political, ideological, social, psychological or other convictions, take a position that is not objective? □ Yes □ No

5.3.1 Please describe the process.

5.4 Do you feel like your integrity and objectivity are sometimes compromised? □ Yes □ No

5.4.1 If yes to question 5.5, please explain further.

5.5 Do you sometimes feel like you taken management roles and responsibilities during audit exercises? □ Yes □ No

5.5.1 If yes to question 5.5, please explain further.

Section F (**Information Assurance**)

6.0 Do users log unto the system using username and password? □ Yes □ No

6.1 Are there any measures to ensure that authorized users have access the data when needed? □ Yes □ No

6.1.1 If yes to question 6.1, please describe the measures.

6.2 Are there any mechanisms in the current system to protect data from been accessed by unauthorized users? □ Yes □ No

6.2.1If yes to question 6.2, please describe the process.

6.3 Are there any mechanisms to ensure data correctness and completeness in the data management process? □ Yes □ No

6.3.1 If yes to question 6.3, please describe the process.

6.4 How are data privacy and confidentiality issues handled in the current system? Please describe the process.

6.5 Describe any other gaps in data management processes if any.

APPENDIX THREE: CLEARANCE CERTIFICATE

R14/49 Ms Mieks Frenken Nyarko Twumasi

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M141164

NAME: (Principal Investigator)	Ms Mieks Frenken Nyarko Twumasi
DEPARTMENT:	School of Public Health VADU Health and Demographic Surveillance System Rual health Program Kem Hospital
PROJECT TITLE:	Standard Operating procedues (SOPs) for Health and Demographic Research Data Quality Assurance, the Case of Vadu Health and Demographic Surveillance System HDSS Site
DATE CONSIDERED:	28/11/2014
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Gideon Nimako and Dr Sanjay K Juvekar
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	04/02/2015

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 FOUR: INFORMATION ASSURANCE POLICY FOR VADU HDSS



Vadu HDSS

INFORMATION ASSURANCE POLICY (IAP)

Terms used in this document

User - Any DMT member authorized to access the Vadu HDSS application.

Privileged Users – Users access levels to the Vadu HDSS application.

Users with full privileges – DM and HDM have full access and are permitted based on job description to add, edit and delete records in a database and make necessary changes to the HDSS system when required.

Users with limited privileges – FRSs, FC and FRAs have limited access to the system due to their job descriptions, from adding, deleting, or changing records in a database.

Virus - a software program capable of reproducing itself and usually capable of causing great harm to files or other programs on the computer it attacks. A true virus cannot spread to another computer without human assistance.

QA – This involve monitoring, auditing, training and documentation. In compliance to GCP guidelines and other required research regulations, it is required of Vadu HDSS DMT to implement and maintain quality assurance and quality control mechanisms with written standard operating procedures to produce quality data. Quality data can be produced in compliance to study protocol by ensuring data is generated, documented, recorded and reported adequately at every stage of data life cycle.

QC- A set of procedures undertaken within the quality assurance system to verify that the requirements for quality of the study protocol have been fulfilled.

Audits: This is designed to evaluate and assure the data consistency and integrity of quality control systems and measure performance against recognized standards. It helps to demonstrate robust research processes.

Training – This is required to ensure the study is in accordance with GCP guidelines and standards. It is a requirement that DMT must undertake GCP training at least once a year.

Anonymized Data - Data relating to an individual where the identifiers have been scrambled or hidden to prevent identification of that individual.

Data - Qualitative or quantitative statements or numbers that are assumed to be factual and not a product of any analysis or interpretation.

Data Sharing – Is the transfer of data from the organization to another organization or to an individual for the purpose of research and/or publication.

Information - Output of some process that summarizes, interprets or otherwise represents data to convey meaning.

1.0 Introduction

1.1 Purpose

This policy outlines the technical controls and security configurations the DMT at Vadu HDSS is required to implement in order to ensure the integrity and availability in data management life cycle. It serves as a central policy document which members of the DMT should be familiar and states actions and prohibitions that should be followed. It provides DMT with guidelines to govern data management life cycle activities using ISACA/COBIT 5 family product and ISO/IEC version 27001 family of ISMS and other standards and regulations. This policy requirements and restrictions defined in this document shall conform to DMP and study protocol. This policy comprises of appropriate training required for team members before the commencement of and during every round, obtaining informed consent, data collection and training, authentication, field monitoring, quality control and assurance, validation checks and cleaning, audit assurance, data security, data storage, archiving and retrieval, data access and sharing. This policy should be adhered to by DMT members to improve the quality of research data and make the process easier for data to be reproduced.

1.2 Scope

This policy covers good practices that will guide the routine activities of DMT at Vadu HDSS in data management life cycle.

2.0 Responsibilities of DMT

The DMT at Vadu HDSS comprises of: HDM, DM, SAD, ITO, FC, FRSs and FRAs.

HDM - The roles and responsibilities of each member of the DMT is assigned by HDM. HDM is responsible for overall planning, designing and execution of all data management software applications and database management in at Vadu HDSS. It is the responsibility of HDM to create new HDSS forms and modify existing forms and also to design database structures based on the physical forms. HDM supervises members of the DMT and ensures the smooth running of data management life cycle activities.

DM - Is responsible for data management activities including running validation checks, range and consistency checks, and quality control checks on data stored in electronic format, cleans data, monitors audit reports, stores and archives data, provides security for stored data and retrieves data when the need arises. DM prepares hardware and software requirements needed for each round. DM is also responsible

for training the FRAs, FRSs and FC to use the HDSS application on the laptops and tablets to achieve quality research data.

SAD - Is responsible for the design of e-forms or the HDSS database or application. SAD ensures that their programs contain the following security features:

- Applications must support authentication of individual DMT members, not groups.
- Applications must not store passwords in clear text or in any easily reversible form.
- Applications must not transmit passwords in clear text over the network.
- Applications must provide for some sort of role management; such that one user can take over the functions of another without having to know the other's password.
- Applications must produce audit logs for adequate monitoring.

FRAs - Is responsible to conduct house to house surveys to enrol study participants, collect and update HDSS data within defined HDSS areas.

FRSs - Supervises FRAs to ensure all households in the HDSS areas in the survey are completed. Ensures that all e-forms are completed without blanks, runs field quality checks and back checks on electronic data. Monitors FRAs to ensure that e-forms are filled correctly. Ensure the smooth running of field work.

FC - Is responsible for planning survey processes and monitoring progress. FC conducts weekly meetings to provide solution to pending field queries.

3.0 General Policy

3.1 Wear Identifying Badge

DMT members should be encouraged to wear their employee identification badge or cards. This is for easy identification and helps build some sort of security.

3.2 Authentication

All computers, laptops, and tablets as well as servers should be secured with usernames and passwords. Every DMT member should be assigned a username and self-created-password, which should be the only way of logging unto Vadu HDSS on computers, laptops and tablets and to have access to the HDSS system. Research data is sensitive and all measures should be put in place to keep it confidential and protect participants' records. Laptops and tablets for instance can easily be stolen, hence logging unto them with usernames and passwords will ensure some level of security in case of they get into the hands of unauthorized persons.

3.3 Password Protection

Passwords are an important aspect of data security. A poorly chosen password may result in unauthorized access and/or abuse of Vadu HDSS's resources. DMT at Vadu HDSS are responsible to take appropriate steps as outlined below, to select and secure their passwords.

3.3.1 Purpose

This is to establish a standard for creating strong passwords, the protection of those passwords and the frequency of change.

3.3.2 Scope

This is to guide members of the DMT at Vadu HDSS in password creation and protection.

3.3.3 Password Creation Policy

All user-level and system-level passwords must conform to the password creation:

DMT members must not use the same password for Vadu HDSS accounts and for other non-Vadu HDSS access (for example, personal password used for say yahoo account)

DMT members must not use the same password for various Vadu HDSS access needs.

DMT members' accounts that have system-level privileges granted through in-depth memberships or programs such as sudo must have a unique password from all other accounts held by that user to access system-level privileges.

3.3.4 Password Change Policy

All system-level passwords (for example, root, enable, NT admin, application administration accounts and others) should be changed on at least a quarterly basis.

All user-level passwords on laptops, tablets and desktops are recommended to be changed at least every six months.

Password cracking or guessing may be performed on a periodic or random basis by ITO. If a password is guessed or cracked during one of these scans, the DMT member will be required to change it to be in compliance with the password creation procedure.

3.3.5 Password Protection Policy

Passwords must not be shared with anyone. All passwords are to be treated as sensitive and confidential as Vadu HDSS data.

All passwords must not be revealed over the phone to anyone.

Do not hint at the format of a password (for example, "my family name").

Do not share Vadu HDSS passwords with anyone including co-workers while on vacation, and family members.

Do not write down and store them anywhere in your office.

Do not store passwords in a file on phones, desktops, laptops and tablets without encryption.

Do not use the "Remember Password" feature of applications (for example, web browsers).

Users suspecting that their passwords may have been compromised must report the incident to ITO and change all passwords.

Please note: Try to create passwords that can easily be remembered but do not write them down. The previous twelve passwords cannot be reused.

3.4 Non-Compliance

Any DMT member found to have violated this policy may be subject to disciplinary action.

3.5 Related Standards, Regulations, Policies and Processes

ISMS, HIPAA, ISO/IEC and ISACA[28] [47] [48]

4.0 Prohibited Activities by DMT

DMT members are prohibited from the following activities:

Modification and/or Configuration Changes - Laptops and tablets owned by Vadu HDSS assigned to users are solely for data management purposes. Modifications or configuration changes are not permitted on these devices for home use.

Personal or Unauthorized Software – The use of personal software is prohibited. All software installed on Vadu HDSS laptops and tablets should be approved by the HDM.

Attempting to break into an information resource or to bypass a security feature - This includes running password-cracking programs or sniffer programs and attempting to avoid file or other resource permissions.

Introducing or attempting to introduce computer viruses, Trojan horses, peer-to-peer (“P2P”) or other malicious code into an information system.

Browsing- It is prohibited to visits certain websites which may introduce viruses of all kinds unto the Vadu HDSS system.

The wilful, unauthorized access or inspection of confidential or sensitive information to which you have not been approved on a "need to know" basis is prohibited.

System Use - Engaging in any activity for any purpose that is illegal or contrary to the policies, procedures or interests of Vadu HDSS are strictly prohibited.

Internet Access - Internet access provided for Vadu HDSS DMT and is considered a great resource for the organization. This resource is costly to operate and maintain and should be allocated primarily to data management activities, administrative and contract needs. The Internet access provided by the Vadu HDSS should not be used for entertainment, listening to music, viewing the sports highlight of the day, games, movies and others. Users must understand that individual internet usage is monitored and if found to be spending an excessive amount of time or consuming large amounts of bandwidth for personal use, disciplinary action will be taken.

Many internet sites, such as games, peer-to-peer file sharing applications, chat rooms, and on-line music sharing applications, have already been blocked by the Vadu HDSS routers and firewalls. This list is constantly monitored and updated as necessary. Users visiting pornographic sites will be disciplined and may have their appointments terminated.

5.0 User Privileges Policy

If a DMT member's position and responsibilities change, the privileges also change. If a DMT member changes positions at the Vadu HDSS, the HDM or ITO shall promptly be notified indicating the new job description and applicable privileges or accesses shall be granted. Every six months, HDM and or ITO shall review and ensure users roles, access, and software necessary to perform their job descriptions effectively while being limited to the minimum necessary data to facilitate ISO/IEC ISMS standards and guidelines compliance and protect Vadu HDSS data.

6.0 Termination of User Logon Account

If a DMT member's appointment is terminated voluntary or involuntary, HDM or ITO shall be promptly notified, where that DMT member's access to the system is removed and the form is submitted to HDM. HDM and/or FRSs shall be responsible for insuring that all Vadu HDSS assets in that DMT member's possession like badges, laptops, tablets and other Vadu HDSS belongings are returned before finally leaving.

7.0 Data Security

7.1 Hardware/Software Security Protections Policy

Antivirus

Vadu HDSS has installed Quick Heal antivirus software on all computers, laptops, tablets and servers. The antivirus software has been scheduled for daily updates to protect data and other software applications. This update is critical to data security, and should be allowed to complete.

Security Locks

Use security cable locks for laptops at all times, even if at home or at the office. Cable locks have been demonstrated as effective in thwarting thefts.

Screens Lock

Users should lock the screen before walking away from the workstation. The data on the screen may be protected to an extent. Ensure the lock feature has been set to automatically turn on after 15 minutes of idleness.

8.2 Backup Policy

Automatic backup should be made available on the laptops, computers and tablets and should be unknown to users.

8.3 Hard copy reports/Paper work Policy

Never leave paper records around your work area. Lock all paper records in a file cabinet at night or when you leave your work area.

8.3.1 Disposal of Paper

All papers which have sensitive information that is no longer needed should be shredded before being disposed. Do not place in a trash container without first shredding. All FRAs, FRSs and FC working from home should bring those papers to the office for shredding.

8.4 Specific Protocols and Devices

8.4.1 Wireless Usage Standards and Policy

Due to an emergence of wireless access points in hotels, airports and in homes, it has become vital that a policy be developed and adopted to ensure the security and functionality of such connections for DMT at Vadu HDSS. Any DMT member who needs wireless usage on laptops and tablets should consult the ITO or HDM at Vadu HDSS.

8.4.2 Software Requirements

DMT members are prohibited to install any software packages except what have already been installed on the device assigned to them based on their roles and responsibilities.

9.0 Data Management Policy

10.0 Informed Consent

10.1 Purpose

Legitimately, effective informed consent should be obtained from every study participant or the study participant's legally authorized representative unless the requirement has been ignored by the IRB in accordance with the ethical principles uttered in the Belmont Report and FDA regulations prior to enrolment and the study. This policy is applicable to all epidemiological, biomedical and other studies conducted by Vadu HDSS which involves human participants. It is a basic prerequisite for Vadu HDSS to obtain both written and oral informed consent from all study participants prior to the study.

10.2 Scope

These processes are requirements applicable to all DMT members designated at the field to obtain informed consent from all study participants.

10.3 Informed Consent Policy

Verbal and written informed consent shall be obtained from all potential study participants before the commencement of every round.

10.4 Non-Compliance

Any member of DMT found to have violated this policy may be subject to disciplinary action.

10.5 Related Standards, Regulations, Policies and Processes

HIPAA[49], The Belmont Report[55], Declaration of Helsinki[43][43](WMA General Assembly, 1964)[2], FDA regulations[13], ICH and GCP Guidelines[56]

11.0 Audit Trails/Controls

To ensure that Vadu HDSS implements hardware, software, and/or procedural mechanisms that record and examine activity in information systems that contain electronic protected research data. Audit controls are technical mechanisms that track and record activities on all workstations including laptops and tablets. An audit trail determines if a security violation occurred by providing a chronological series of logged computer events that relate to an operating system, an application, or user activities based on protocols. It is required that the ITO/DM must constantly audit DMT members' activities in order to continually assess potential risks and vulnerabilities to research data in their possession and develop, implement and maintain appropriate administrative, physical and technical security measures in accordance with the HIPAA Security Rule.

11.1 Purpose

This outlines the policy put in place to guide DMT at Vadu HDSS to ensure that DMP and study protocol are followed to produce quality data through monitoring and audit trail recordings.

11.2 Scope

All members of the DMT at Vadu HDSS should be familiar and adhere to this policy while performing their routine activities.

11.3 Audit Trail policy

These functions should be recorded:

- log-in attempts
- password changes,
- file creations, changes and/or deletions to the system.
- The audit trail event record should specify:
- type of event,
- when the event occurred – recording date, time,
- user ID associated with the event, and
- Program or command used to initiate the event.
- Create a track of data corrections

11.4 Non-Compliance

Any DMT member found to have violated this policy may be subject to disciplinary action.

11.5 Related Standards, Regulations, Policies and Processes

ISACA[57][50][25]

12.0 Training

The primary goal of Vadu HDSS is to improve and/or produce quality research data that could enable policy makers make appropriate decisions to positively affect Vadu community as a whole. To achieve this, DMT at the field (FRAs, FRSs and FC) collecting data should be trained to use the HDSS application on the laptops and/or tablets to capture all required responses from study participants during enrolment and various event updates.

12.1 Purpose

The HDM and/or DM must ensure all members understand the entire data collection process, verify the accuracy and completeness of data in compliance to study protocol and DMP.

12.2 Scope

The training is organized for all members of the DMT at Vadu HDSS working on the field.

12.3 Training Policy

FC, FRSs and FRAs should be trained to understand and be familiar with the HDSS application on laptops and/tablets to capture required responses from study participants during enrolment and various events updates adequately while ensuring data confidentiality, integrity and security at the beginning of each round.

12.4 Non-Compliance

Any member of DMT at the field found to have violated this policy may not be eligible to be a member of the DMT at the field.

12.5 Related Standards, Regulations, Policies and Processes

SANS, ISACA[57], GCP guidelines[51], [58], [(59]

13.0 Field Monitoring

To ensure that FRAs are capturing data correctly in accordance with Vadu HDSS DM plan and study protocol. FC and/or FRSs should physically visit the field to monitor FRAs' work. The time of field visit is unknown to FRAs, this is to ensure that FRAs is not cooking data but actually capturing the right responses from study participants. SAD has incorporated an electronic monitoring mechanism on the HDSS applications on the tablets.

13.1 Purpose

Field monitoring are mechanisms to ensure that FRAs are at their designated locations and capturing correct responses from study participants.

13.2 Scope

This applies to all FC, FRSs, SAD and DM to ensure that all FRAs are tracked while conducting their routine activities on the field.

13.3 Monitoring Policy

Routine data collection should be monitored by FC and FRSs physically and electronically by the HDSS application on the laptops and/or tablets.

13.4 Non-Compliance

Any member of DMT found to have violated this policy may be subject to disciplinary action.

13.5 Related Standards, Regulations, Policies and Processes

ISACA standards[60],[28] and FDA regulations[21],[47],[48], [49],[13]

14.0 QC/QA

This is a crucial component of a data management. It provides the basic knowledge required to accomplish a procedure correctly. Training also provides the understanding of a given task or procedure, thereby enabling DMT involved to make informed and effective decision.

14.1 Purpose

This is to ensure that the study is performed and the data generated, documented and reported are in compliance with GCP and the applicable regulatory requirements. All members of the DMT should adhere to this document to produce quality data.

14.2 Scope

This policy applies to all members of DMT at Vadu HDSS.

14.3 QC/QA Policy

Vadu HDSS DMT must implement all QC/QA at all stages of data management life cycle to ensure quality research data at the end of the study in accordance with DMP and study protocol.

14.4 Non-Compliance

Any member of DMT found to have violated this policy may be subject to disciplinary action.

14.5 Related Standards, Regulations, Policies and Processes

ISMS [61], ISACA [62], GCP[52] and [60]

15.0 Query Resolution

15.1 Purpose

DM, FC and/or FRSs must ensure that all queries are resolved in accordance with DMP and study protocol. This is to reduce errors during data capture. This is to promote consistent, efficient and effective data management life cycle.

15.2 Scope

All FRAs, FRSs, and FC should be familiar with this policy while performing all routine activities.

15.3 Query Resolution Policy

DMT should be trained to understand the proper methods of resolving queries in accordance with DMP and study protocol as well as internationally acceptable guidelines and standards.

15.4 Non-Compliance

Any member of DMT found to have violated this policy may be subject to disciplinary action.

15.5 Related Standards, Regulations, Policies and Processes

ISMS [61] , [56] and GCP[52]

16.0 Validation Checks and Cleaning

16.1 Purpose

HDM, SAD and/or DM at Vadu HDSS have designed a validation check list to help reduce errors, ensure accuracy and completeness of data on all e-form types on the HDSS application in accordance to DMP and study protocol.

16.2 Scope

This document applies to all members of the DMT at Vadu HDSS.

16.3 Validation Policy

SAD and/or DM must run the program on the data weekly in accordance with Vadu HDSS DMP validation checklist.

16.4 Non-Compliance

Any member of DMT found to have violated this policy may be subject to disciplinary action.

16.5 Related Standards, Regulations, Policies and Processes

ISACA[9], COBIT [16] , and GCP[52]

17.0 Data Storage and Archiving Policy

Vadu HDSS have a number of backup and storage mechanisms. These mechanisms are put in place to prevent loss of data in case of any system failure after data cleaning and secure data for future retrieval at the end of every round.

17.1 Purpose

This is to secure data for future retrieval at the end of every round in case of any system failure after data cleaning.

17.2 Scope

All members of the DMT should adhere to this policy to secure data for easy retrieval.

17.3 Data Storage Policy

Clean data should be saved and updated weekly on all servers (local and cloud servers) and storage devices available for DMT at Vadu HDSS.

17.4 Related Standards, Regulations, Policies and Processes

ISMS[48], Data archiving[53][46], GCP[52] and CDISC[42]

18.0 Data Access and Sharing

18.1 Purpose

Data collected using public and charitable funds must show the untapped potentials of research and benefits to social society, and it is essential to make available non-sensitive data for legitimate and registered use.

The data is generated and compiled for specific requirements. Data generated for different purposes have different structures and formats and are not stored in the same storages giving rise to the issues of standardized format and inter-operability of both scientific and technical nature. Data collected from the population should be responsibly shared with global researchers to produce quality findings thus helping in formulation of useful rural health policies while ensuring participants' data privacy and confidentiality. Data is anonymized to hide study participants' identifiers. Then datasets are labelled with standard identifiers and versions so that users can easily distinguish and compare separate analyses of datasets. Researchers should be able to link associated datasets stored in various databases and link datasets to any publications based on the data.

18.2 Scope

The HDM and DM at Vadu HDSS should adhere to this policy for data access and sharing with the public.

18.3 Data Access and Sharing Policy

Vadu HDSS have designed this policy to promote data sharing and enable authentic and registered access to its electronically stored data for research and publications. It is expected that before publication, the data are accessed under the data sharing and access guidelines.

18.4 Related Standards, Regulations, Policies and Processes

GCP[63] , ISACA[62], and KEMHDSS Pune Data Access and Sharing Policy [54].

APPENDIX FIVE: STANDARD OPERATING PROCEDURES (SOPs)

VADU RURAL HEALTH PROGRAM
KEM Hospital Research Centre, Pune

STANDARD OPERATING PROCEDURE (SOP)

For

Vadu HDSS

Data Management Life Cycle

Review: Six Months

Developed by: Mieks Frenken Nyarko Twumasi
Student, University of the Witwatersrand

Supervised by:

External Supervisor: Dr. Sanjay K. Juvekar
Officer-In-Charge, Vadu HDSS

Prof. Tathagata Bhattacharjee
Head of Data Management, Vadu HDSS

Internal Supervisor: Mr. Gideon Nimako, Lecturer and Program Coordinator,
MSc. Epidemiology in Research Data Management, Wits

Sponsored by: INDEPTH Networks, Ghana

Version: V01_1.0.0

Date: 18/06/2016

History of confirmed version

Version Number Date dd/mmm/yyyy	Version pages	Description of change	Approved By

Content of the SOP

1.0 Scope
2.0 Acronyms
3.0 Terms used in this document
4.0 Guidelines
  4.1 Authentication
  4.2 Inform Consent Process
  4.3 Audit Trails
  4.4 Training
  4.5 Field Monitoring
  4.6 Quality Control and Quality Assurance
  4.7 Query Resolution
  4.8 Validation Checks
  4.9 Data Storage and Archiving
  4.10 Data Access and Sharing.
5.0 Related documents**,** standards, regulations, policies and processes

1.0 Scope

This is to guide DMT members at Vadu HDSS to perform their routine activities. All members of the DMT should adhere to this document.

2.0 Acronyms

DM	Data Manager
DMT	Data Management Team
DMP	Data Management Plan
FRA	Field Research Assistant
FS	Field Supervisor
FC	Field Coordinator
HDM	Head of Data Management
ITO	Information Technology Officer
QA	Quality Assurance
QC	Quality Control
SAD	Software Application Developer

3.0 Terms used in this document

User - Any DMT member authorized to access the Vadu HDSS application.

Privileged Users – Users access levels to the Vadu HDSS application.

Users with full privileges – DM and HDM have full access and are permitted based on job description to add, edit and delete records in a database and make necessary changes to the HDSS system when required.

Users with limited privileges – FRSs, FC and FRAs have limited access to the system due to their job descriptions, from adding, deleting, or changing records in a database.

Virus - a software program capable of reproducing itself and usually capable of causing great harm to files or other programs on the computer it attacks. A true virus cannot spread to another computer without human assistance.

QA – This involve monitoring, auditing, training and documentation. In compliance to GCP guidelines and other required research regulations, it is required of Vadu HDSS DMT to implement and maintain quality assurance and quality control mechanisms with written standard operating procedures to produce quality data. Quality data can be produced in compliance to study protocol by ensuring data is generated, documented, recorded and reported adequately at every stage of data life cycle.

QC- A set of procedures undertaken within the quality assurance system to verify that the requirements for quality of the study protocol have been fulfilled.

Audits: This is designed to evaluate and assure the data consistency and integrity of quality control systems and measure performance against recognized standards. It helps to demonstrate robust research processes.

Training – This is required to ensure the study is in accordance with GCP guidelines and standards. It is a requirement that DMT must undertake GCP training at least once a year.

Anonymized Data - Data relating to an individual where the identifiers have been scrambled or hidden to prevent identification of that individual.

Data - Qualitative or quantitative statements or numbers that are assumed to be factual and not a product of any analysis or interpretation.

Data Sharing – Is the transfer of data from the organization to another organization or to an individual for the purpose of research and/or publication.

Information - Output of some process that summarizes, interprets or otherwise represents data to convey meaning.

Below are the procedures found in the IAP document for to guide data management life cycle:

4.0 Guidelines/steps/procedures

4.1 Authentication - Password Creation, Change and Protection

4.1.1 Purpose

The procedure is to establish a standard for creating strong passwords, the protection of those passwords and the frequency of change.

4.1.2 Password Creation Guidelines

The following are the guidelines to create strong password:

To gain assess

- All passwords should meet or exceed the following guidelines and procedures.
- Strong passwords have the following characteristics:
- Contain at least 12 alphanumeric characters.
- Contain both upper and lower case letters.
- Contain at least one number (for example, 0-9).
- Contain at least one special character (for example,!$%^&*_+=\`:";'<>?,/).

Poor, or weak, passwords have the following characteristics:

- Contain less than eight characters.
- Can be found in a dictionary, including foreign language, or exist in a jargon.
- Contain personal information such as birthdates, addresses, phone numbers, or names of pets, friends, and fantasy characters.
- Contain number patterns such as aaabbb, qwerty, zyxwvuts, or 123321.
- Contain common words spelled backward, or preceded or followed by a number (for example, secret1 or 1secret). Are some version of “Welcome123”, “Password123”, “Changeme123”

Please note: Try to create passwords that can easily be remembered but do not write them down. The previous twelve passwords cannot be reused [28][47][48].

4.2 Informed Consent Process

4.2.1 Purpose

This outlines the processes for obtaining written informed consent for enrolling participant into the HDSS study. This is in accordance to The Declaration of Helsinki and ICH GCP that it is necessary for DMT members at the field to ensure that the potential participants understand what they are undertaking when they sign a consent form for research by means of a written Participant Information Sheet and a verbal explanation in the form of a study discussion or talk.

4.2.2 Guidelines involved in obtaining consent from study participants

- **Obtain Permission** - Permission must be granted by household head before obtaining informed consent from potential study participants.
- **Purpose of research** – FRAs, FRSs and FC must give details of the study, purpose of the study, expected duration of participation and the procedures that will be followed.

- **Risk and Benefits** - Any anticipated risks and benefits to participants must be explained in details to potential study participants.
- **Possible Treatment** - Any available courses of treatment that might be beneficial to the participants must be disclosed.
- **Privacy and Confidentiality** – Potential study participants must be assured that their records shall be kept private and confidential.
- **Compensation/Reimbursement** - Any compensation made available to potential study participants must be disclosed as a result of either participation, or any medical treatments in case of injury.
- **Voluntary Participation and Withdrawal** - Study participants must understand that participation is voluntary, refusal to participate or withdrawal in the course of the study will not involve any penalty or loss of benefits.
- **Disclosure of results** – Potential study participants must be assured that the results of the study would be related to them when available.
- **Written Informed Consent** - Informed consent must be written in the dialect well understood by study participants to enable them make informed decisions whether or not to participate in the study.
- **Fill Informed Consent Form**- Informed consent form must be filled and signed or thumb printed by study participants. Participants signature or thumb print can be captured electronically on the HDSS application available on the tablets [28] [43] [55]

4.3 Audit Trail

4.3.1 Purpose

This outlines the processes put in place by DMT at Vadu HDSS to ensure that DMP and study protocol are followed to produce quality data through monitoring and audit trail recordings.

4.3.2 Instructions of incorporating audit trail

- SAD and/or DM have incorporated some mechanisms of system online monitoring and audit trail recording, protecting, reviewing and reporting security breaches or anomalies to the HDM.
- SAD and or DM periodically monitor online programmer activity to ensure audit trail functions are operating and reports are reviewed weekly and it should be able to aid SAD and/or DM to reconcile audit trail anomalies.

- ITO/DM shall enable event auditing on all computers that process, transmit, and/or store research data for the purposes of generating audit logs. Each audit log shall include: user ID, login time and date, details of data being accessed for each attempted access. Audit trails shall be stored on separate workstations to reduce the impact of audits trails or logs been accessed on individual laptops and tablets.
- HDSS audit files shall be stored in a locked room and kept according to protocol.
- HDM/ITO/DM/SAD shall be responsible for monitoring audit trails to check if there are protocol violations and any inappropriate validations [2] [13] [55].

4.4 Training

4.4.1 Purpose

For Vadu HDSS to improve and/or produce quality research data, DMT at the field (FRAs, FRSs and FC) collecting data must be trained to use the HDSS application on the laptops and/or tablets to capture all required responses from study participants during enrolment of participants and various event updates. The HDM and/or DM must ensure all members understand the entire data collection process, verify the accuracy and completeness of data in compliance to study protocol and DMP.

4.4.2 Training Steps are as follows:

- HDM and/or DM must organize training sessions for all DMT (FC, FRSs and FRA) on the field to understand how the HDSS application works.
- FC and FRSs must ensure that all FRAs have made themselves available for the training sessions at the beginning of each round and subsequent training sessions.
- Each member of the DMT responsible for collecting data must be fully trained on the HDSS application on both laptops and tablets prior to the beginning of the study.
- Each member of DMT will have proper security privileges assigned prior to entering data into the HDSS application.
- No member of DMT with security privileges will grant access to another person under their identity and password.
- Only members of DMT with proper security access will have access to the HDSS application.
- Different privileges shall be granted to FC, FRSs and FRAs.
- A paper or oral based assessment should be conducted to ascertain the level of understanding of the training undertaken[57][50][25].

4.5 Field Monitoring

4.5.1 Purpose

To ensure that FRAs are capturing data correctly in accordance with Vadu HDSS DMP and study protocol. FC and/or FRSs must physically visit the field to monitor FRAs' work. The field visit is unknown to FRAs, this is to ensure that FRAs are not cooking data but actually capturing the right response from study participants. SAD has incorporated an electronic monitoring mechanism on the HDSS applications on the tablets.

4.5.2 Field Monitoring Guidelines are as follows:

- **Work Plan** - FRAs must design a weekly work plan and submit to FRSs and/or FC.
- **Daily monitoring** - With the weekly work plan, FRSs and/or FC knows the location where each FRA is. 12 FRAs are assigned to one FRSs. FRSs must visit about 60% of FRAs in a week.
- **Phone calls** - FRSs and/or FC calls the FRAs in case they are not found in the expected location for direction because some households are far apart.
- **Cross-Checks Response Captured** - FRSs and/or FC randomly cross-checks responses captured on the HDSS application with study participants to ensure whether data captured is accurate.
- **Record Interview** - HDSS applications on the tablets have been programmed to make a voice recording of the interactions between FRAs and study participants during data collection but the voice recording is unknown to FRAs.
- **GPS** - GPS application on tablets have been activated and programmed to capture location where the interviews and data collection was done but unknown to FRAs.
- **FC Monitoring** - FC visit field to ensure smooth running of all field work with accordance to DMP and study protocol [51][57].

4.6 Quality Control(QC)/Quality Assurance (QA)

4.6.1 Purpose

This outlines the QC and QA measures put in place for data management life cycle at Vadu HDSS. This is to ensure that the study is performed and the data generated, documented and reported in compliance with GCP and the applicable regulatory requirements. All members of the DMT should adhere to this document to produce quality data.

4.6.2 QC/QA Steps as follows:

- HDM and/or DM must ensure that all required QC/QA processes are followed in accordance with HDSS study protocol and validation checklist of DMP. This involves checking protocol with electronic data by ensuring that:
    - All validation checks are run on all captured data.
    - If more errors are detected after validation checks by some FRAs who captured data, training sessions are organized for them by DM.
- The HDSS application on the laptops and tablets has been programmed to capture all required responses else the e-form cannot be saved. The HDSS application would not be saved if it detects the following:
    - Protocol violation
    - Missing values
    - Outliers (Range checks)
    - Inconsistencies and others [60][28][21][47][48][49][13]
    -

4.7 Query Resolution

4.7.1 Purpose

DM, FC and/or FRSs must ensure that all queries are resolved in accordance with DMP and study protocol. This is to promote consistent, efficient and effective data management life cycle.

4.7.2 Query Resolution Guidelines

DMT must be trained to resolve queries in accordance with the following requirements:

- Checks daily for queries.
    - All queries must be resolved within 2 weeks unless specified by study protocol.
    - If more information is required to resolve queries, the study protocol should be referred.
- Study participants will have to be consulted for queries with exceptional problems.
- DM will reference to unanswered query list with FC/FRSs for outstanding queries weekly [61][56][52].

4.8 Validation Checks and Cleaning

4.8.1 Purpose

HDM, SAD and/or DM at Vadu HDSS have designed a validation check list to help reduce errors, ensure accuracy and completeness of data on all e-form types on the HDSS application in accordance to DMP and study protocol.

4.8.2 Validation Guidelines

The validation checklist ensures the following guidelines:

Validation Checks

- Every DMT member must log unto the HDSS systems using assigned username and password.
- Every laptop and/or tablet is assigned to only one field worker.
- FRAs are encouraged to keep their passwords private and secret.
- System generates date and time – The HDSS application generates date and time automatically like the following:
    - **Interview Start Time**
        - Time in Hours (06:00am to 21:00pm) and Minutes
        - Start Time should be less than End Time
        - Time format should be 24 hours.
    - **Interview Date**
        - Mandatory (Not Null)
        - Interview Date should not be Future date
        - Interview Date should be greater than date of Birth.
    - **Field worker name**
        - Mandatory (Not Null)
        - The system automatically displays field worker's name.
        - Interview End Time
        - Mandatory (Not Null)
        - Time in Hours (06:00am to 21:00pm) and Minutes.
        - End Time should be greater than Start Time
        - Time format should be 24 hours.
- **Ensuring accuracy and completeness**
    - The HDSS application will not save information captured if FRAs does not fill all required answers to questions.
    - FRAs must select appropriate values from the options provided on the HDSS application which helps to reduces errors.
- **Privileges**
    - Limited Access - FRA can only view, edit and update entry screen on HDSS application but cannot delete responses entered earlier.

- Full Access – HDM/DM have full access to the system. They can view, edit, update and delete.

- **Query Resolution** - The HDSS application prompts DM for outstanding queries which must be resolve.
- **Field QC and Back Checks** - FRSs runs the field quality control checks and back checks done on data captured.
- **Final Checks** - After query resolutions, validation checks are executed finally to ensure data is cleaned.
- **Audit Checks** - If DM detects that responses captured have many errors, or responses are captured less than an average time, then DM suspects FRAs for improper evaluations and FRAs are questioned for explanation.

4.8.3 Protocol Reading

HDM and/or DM must read and understand study protocol to ensure protocol is in conformity with DMP and validation checklist. This is required for appropriate updates to the HDSS application [9][16][52].

4.9 Data Storage and Archiving

4.9.1 Purpose

These are measures put in place to protect data at Vadu HDSS. These measures are backup and storage mechanisms. These are put in place to secure data for future retrieval at the end of every round in case of any system failure after data cleaning.

Data Storage and Archiving Guidelines

- FRAs are required to upload the responses captured on the HDSS system at the close of work daily to the server without necessarily physically present at the office if there is the availability of network.
- The clean data on the local and cloud server is updated weekly.
- Copies of clean data are uploaded unto every individual laptop and tablet for the next week field visits.
- Reports are generated weekly by the FC aided by DM[48][53][55]

4.10 Data Access and Sharing

4.10.1 Purpose

This outlines the processes DMT at Vadu HDSS have in place for data to be accessed and shared with the public. Large volumes of data are generated and compiled for specific

requirements, data generated for different purposes have different structures and formats; and are not stored in the same storages giving rise to the issues of standardized format and inter-operability of both scientific and technical nature. While ensuring the privacy and confidentiality of participants, data collected from the population must be shared responsibly with global researchers to produce quality findings to aid in formulation of useful policies. Data is anonymized to hide study participants' identifiers. Datasets are labelled with standard identifiers and versions so that users can easily differentiate and compare separate analyses of datasets. Researchers should be able to link associated datasets stored in various databases and link datasets to any publications based on the data.

4.10.2 Data Access and Sharing Guidelines

- **Data Sharing Stages**

  This involves three guidelines, which is as follows:

  - Primary Data - Data can be shared immediately after basic Quality Control (QC) checks and DMT members are clearly notified of the level of QC which has taken place. This is shared amongst the internal teams or departments.
  - Intermediate Results - The results of intermediate analysis can be shared with the internal teams, other departments, Principal Investigators of linked projects or external evaluators for verifying the processes and results.
  - Final datasets - This dataset is ready for sharing and can be shared based on the access levels. The DMT members can submit their versions of the final datasets with details regarding the OPERATINGs carried out by them on the datasets.

- **Data Access levels**

  The access level assigned to data will guide data owners, data custodians, DMT members, technical teams and any others who may obtain or store data, to implement the security protections and access authorization mechanisms appropriate for that data. Such categorization encourages the discussion and subsequent full understanding of the nature of the data being displayed or manipulated.

  - **Public (low level of sensitivity)** - Access to "Public" data may be granted to any requester. Public data is not considered confidential. The integrity of public data must be protected and it cannot be released (copied or replicated) without appropriate approvals. All usages must be acknowledged.
  - **Restricted** - Access to "Restricted" data must be controlled from creation to destruction, and will be granted access to only those persons whose requests are

formal and approved by the competent or designated authority. All usages must be acknowledged.

- **Sensitive** - Access to “Sensitive” data must be requested from, and authorized by, the Data Owner. Data may be accessed by persons as part of their job responsibilities. The integrity of this data is of primary importance, and the confidentiality of this data must be protected. An example of Sensitive data includes un-anonymized datasets. All usages must be acknowledged.

- **Data Storage Structure**

  There will be a directory for each dataset. The main dataset will be stored at the top level in the directory. User-submitted datasets corresponding to a dataset will be stored in the directory containing the original dataset. The directory structure for storing the user-submitted datasets will be created using key-word and possible second level key-words. The key-words will be suggested by the data owners and/or the DMT members submitting the datasets. Types of DMT members:

  - **Normal user** - Any user who wants to use the data becomes a normal user. S/he will fill up a form to request a dataset.
  - **Site Administrator** - This user will have all the access to all the data storage. This user can grant access to the requests received from the normal DMT members after following guidelines given in this document and checking with the proper authorities. The data will be shared using NADA system. All the data requests will be sent to the site administrator. The site administrator can give access to download a dataset to a user. If any communication is received regarding the decision for not granting access, then the administrator will forward this communication to the PI of the concerning project within a week of receiving such communication [63][62][54].

5.0 Related Standards, Regulations, Policies and Processes

ISMS, HIPAA,ISO/IEC,ISACA, The Belmont Report, Declaration of Helsinki, (WMA General Assembly, 1964), FDA regulations, ICH and GCP Guidelines, ISACA standard, and KEMHDSS Pune Data Access and Sharing Policy