Research Data Management in Kenya's Agricultural Research Institutes

Emily Jeruto Ng'eno

Bsc (Hons.), MBibl

Submitted in fulfillment of the requirements for the degree of Doctor of Philosophy (Information Studies) in the School of Social Sciences, College of Humanities, University of KwaZulu-Natal, Pietermaritzburg - South Africa

Supervisor:

Prof. Stephen Mutula

March, 2018

DECLARATION

I, Emily Jeruto Ng'eno, declare that:

a) This thesis, except where otherwise indicated, is my original research;
b) This thesis has not been submitted for any degree or examination at any other university;
c) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;
d) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   i. Their words have been re-written but the general information attributed to them has been referenced.
   ii. Where their exact words have been used, then their writing has been placed in single spacing and indented, and referenced.
   iii. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References section.

Student Name: Emily Jeruto Ng'eno
Date:
Supervisor: Prof. Stephen Mutula
Date:
Signature:

ABSTRACT

Research Data Management (RDM) refers to the collection, organization, validation, and preservation of data for analysis, discovery, sharing, reuse and transformation. RDM consists of a number of different activities and processes that include creation of data, storage, security, preservation, retrieval, sharing, and reuse while taking into account technical capabilities, human resource capability, ethical considerations, legal issues and government. The strategic importance of RDM within agricultural research institutes is to: enable scrutiny of research findings, prevent duplication of effort by enabling others to use the same data; promote innovation through retrieval, co-analysis of data, ensuring research data gathered is not lost or destroyed, and that the research meet funders' requirements.

The purpose of this study was to examine Research Data Management (RDM) in Kenya's agricultural research institutes with the view to proposing interventions to improve management, sharing and reuse of agricultural research output. The objectives of the study were to: 1) assess the status of research data management in Kenya's agricultural research institutes; and 2) to determine the legal and policy framework, ICT infrastructure and human capital that is available to facilitate RDM in Kenya's agricultural research institutes.

The study was underpinned by the Community Capability Model (CCM) framework (Lyon, Ball, Duke and Day, 2012) and Data Curation Centre (DCC) Lifecycle Model (Higgins, 2008). The study adopted pragmatism ontology with mixed methods epistemology that enabled the researcher to collect quantitative data from a large sample of researchers in six purposively selected research institutes. Census was used to select the respondents who consisted of directors of institutes, heads of research, heads of IT and librarians. Both quantitative and qualitative data were collected. Quantitative data was analyzed using SPSS to generate descriptive and inferential statistics while the qualitative data was analyzed thematically.

The findings of the study revealed that RDM legal framework did not exist in the institutes surveyed; the RDM policies and regulations were outdated; the institutes lacked unit/department to coordinate functions of RDM; there was limited RDM awareness and advocacy; the institutes lacked RDM security systems; the institutes suffered from lack of or inadequate RDM guidelines on standardization; technical infrastructure; skills and collaborative partnerships. Overall, the findings revealed that RDM was poorly managed. The study recommended among others, the establishment of a formal data governance structure to

address RDM issues, a legislative and policy framework for RDM; capacity building programs and plans, incentivisation of researchers; and a sound technical infrastructure.

ACKNOWLEDGEMENT

First, I want to thank the almighty God for his favor, grace, mercy and knowledge which He granted me during my PhD program. I give Him the initial glory and honor. I wish to thank the following individuals and organizations who, in one way or another, made it possible for the successful completion of this study.

I am indebted to my supervisor Prof. Stephen Mutula for his admirable devotion to my studies, close supervision and technical advice in overcoming obstacles faced during the writing of my thesis. In his own right, Prof. Mutula is a gifted counselor, exceptional supervisor and above all an extraordinary teacher. He gave selflessly of himself and his time without reservation to help me with this PhD Journey. THANK YOU PROF.

My gratitude goes to my employer, Moi University, for granting me study leave to undertake this PhD program. I extend my sincere thanks to the Director Kenya Agricultural and Livestock Research Organization (KALRO) for granting me permission to undertaken research in the six research institutes under KALRO. I am equality grateful to the six Institute directors and their staff for their immense contribution in providing data and information required during my research period.

My special thanks go to my husband and friend Dr Tuitoek Kipruto and our children Ethan, Eliora and Eliana for their unwavering and unconditional physical, emotional and spiritual support throughout my study. My sincere and heartfelt thanks go to my dad Mr. Alfred Ng'eno for introducing me to the world of education and sacrificing every bit of his comfort to see me through in the world of academia. I do appreciate my sister Janet for supporting and standing by my family while away-'*Kongoi missing'*. To the larger family, brothers and sisters for their continuous prayers, support and encouragement- '*Ahsanteni sana'.*

I wish to acknowledge Mr. Barno for his assistance during my research period and especially securing the gatekeepers letter. Further acknowledgement goes to Beatrice, Rehema, Robert, Nokuphila and Amoth for their enormous support morally and academically during my studies.

All persons who made contributions in this endeavor, one way or the other are acknowledged– THANK YOU ALL. GOD BLESS

DEDICATION

'I truly dedicate this thesis to God'

"Praise the Lord, O my soul; Lord my God, you are great; I give thanks to the Lord"

Psalm 104:1

"I will praise you, Lord my God, with all my heart; I will glorify your name forever"

Psalm 86:12

I can do all things through Christ who strengthens me.

I also dedicate this thesis to research institutions.

TABLE OF CONTENTS

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACRL	: Association of College and Research Libraries
AIMS	: The Agricultural Information Management Standards
ANDS	: Australian National Data Services
ARC	: Australian Research Council
ASCU	: Agricultural Sector Coordination Unit
BRI	: Biotechnology Research Institute
CABI	: Centre for Agriculture and Biosciences International
CCM	: Community Capability Model
CGIAR	: Consultative Group for International Agricultural Research
CIARD	: Coherence in Information for Agricultural Research for Development
CIRAD	: The Centre for International Cooperation in Agricultural Research for Development (France)
CRI	: Coffee Research Institute
CSIR	: Council for Scientific and Industrial Research
DCC	: Data Curation Centre
DFID	: Department for International Development (United Kingdom)
DIRISA	: Data Intensive Research Initiative of South Africa
DRI	: Dairy Research Institute
ESRC	: Economic and Social Research Council
EPSRC	: Engineering and Physical Sciences Research Council
FAO	: Food and Agriculture Organization
FCRI	: Food Crop Research Institute
GDP	: Gross Domestic Product
GeRRI	: Genetic Resources Research Institute
GFAR	: Global Forum for Agricultural Research
GIS	: Geographic Information Systems
ICIPE	: The International Centre of Insect Physiology and Ecology
ICPSR	: Inter-University Consortium for Political and Social Research
ICRAF	: World Agroforestry Centre
ICT(s)	: Information and Communication Technologies

IFAD	: The International Fund for Agricultural Development
ILRI	: International Livestock Research Institute
ISO	: International Standards Organization
IT	: Information Technology
IWG	: Interagency Working Group
JISC	: Joint Information Systems Committee
JKUAT	: Jomo Kenyatta University of Agriculture and Technology
KAINet	: Kenya Agricultural Information Network
KALRO	: Kenya Agricultural and Livestock Research Organization
KARI	: Kenya Agricultural Research Institute
KARI-NARL	: Kenya Agricultural Research Institute-National Agricultural Research Laboratories
KEFRI	: Kenya Forestry Research Institute
KEMRI	: Kenya Medical Research Institute
KIRDI	: Kenya Industrial Research Institute
MoA	: Ministry of Agriculture
MOU	: Memorandum of Understanding
NACOSTI	: National Commission for Science, Technology and Innovation
NARA	: National Archives and Records Administration
NARS	: National Agricultural Research System
NCRIS	: National Collaborative Research Infrastructure Strategy
NCSU	: North Carolina State University
NDSA	: National Digital Stewardship Alliance
NeDICC	: Network of Data and Information Curation Communities
NERC	: Natural Environment Research Council
NHMRC	: National Health and Medical Research Council
NICIS	: National Integrated Cyberinfrastructure System
NIH	: National Institutes of Health
NRF	: National Research Foundation
NSF	: National Science Foundation
NSTCC	: National Science and Technology Council Committee

OAIS	: Open Archival Information System
OECD	: Organization for Economic Co-operation and Development
RDM	: Research Data Management
RECODE	: Recommendations for Open Access to Research Data in Europe
REF	: Research Excellence Framework
SAS	: Statistical Analysis System
SCONUL	: The Society of College, National and University Libraries
SIM4RDM	: Support Infrastructure Models for Research Data Management
SOPs	: Standard Operating Procedures
SPSS	: Statistical Package for the Social Sciences
SSHRC	: Social Sciences and Humanities Research Council
STFC	: Science and Technology Facilities Council
TRI	: Tea Research Institute
UCT	: University of Cape Town
UK	: United Kingdom
UKZN	: University of KwaZulu-Natal
UKOLN	: United Kingdom Office for Library and Information Networking
UNCST	: The Uganda National Council of Science and Technology
UNDP	: United Nation Development Programme
UNEP	: United Nations Environment Programme
UNISA	: University of South Africa
USA	: United States of America
WFP	: The World Food Programme

CHAPTER ONE

INTRODUCTION AND BACKGROUND TO THE STUDY

1.1 Introduction

The purpose of this study is to examine Research Data Management (RDM) practices in Kenya's agricultural research institutes with the view to proposing interventions to improve management, sharing and reuse of agricultural research output. Research data are valuable resources that need to be managed by research institutes as they are the original sources or material that researcher(s) have created or collated in conducting a research project of which the research data can be in digital or non-digital form (Ray, 2014). Dora and Kumar (2015) note that research data can be presented in three forms: raw data directly produced from a laboratory or survey; processed data which has been cleaned, refined, arranged and combined in a manner that is useful in research; and data published in journals, in other alternate materials of a similar nature or in other scientific communication. Research data being intricate and complex, they are irreplaceable, expensive and time-consuming to replicate however, there is need for research institutes to be accurate and precise with their collection, description, preservation techniques, access, reuse and sharing of research data (Fellous-Sigrist, 2015; University of California, Los Angeles, n.d).

The term research data has been defined by different scholars in different ways. Boston University Libraries (n.d) defines research data as data that is collected, observed, or created, for the purpose of analysis to produce original research results in forms of raw data, abstracted or analyzed, experimental or observational which include but are not limited to laboratory notebooks, field notebooks, questionnaires, audiotapes, videotapes, photographs, specimens, samples artifacts, among others. Meanwhile, North Carolina State University (NCSU) Libraries (n.d) defines research data as recorded factual material commonly accepted in the scientific community as necessary to validate research findings which are intangible (as in measured numerical values) or tangible (as in physical research materials) and can be generated or collected from simulations, interviews, observations, surveys, experiments, or even from previous literature. From the two definitions it is clear that research data are gathered through a variety of methods including experimentation, observation, interviews, survey, and repurposing of existing data which are captured, described, preserved, accessed for use and reuse, and sharing in order to increase efficiency, safety, quality, reputation and compliance of research data.

Studies show that when agricultural research data is properly managed and shared, it enables researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographical and disciplinary boundaries (Government of Canada, 2016; Alila and Atieno, 2006; Chisenga, Kedemi, Sam, Rudgard and Martin, 2011; and Higman and Pinfield, 2015). Intrinsically, agricultural research data serves an important purpose in generating valuable information and knowledge that is vital in agricultural research institutes and plays a pivotal role in enhancing agricultural productivity and addressing the problem of food insecurity. The Government of Canada (2016) maintains that the ability to preserve, access, reuse and build upon research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for productivity, competitiveness, and quality of life. Therefore, there is a need for agricultural research institutes to engage in RDM.

Research data management (RDM) on the other hand has also been defined in various ways by different scholars. Ray (2014) defines Research Data Management (RDM) as the collection, organization, validation, and preservation of data for analysis, discovery, sharing, reuse, and transformation. Whyte and Tedds (2011) define RDM as the organization of data from its entry into the research cycle through to the dissemination and archiving of valuable results aiming to ensure reliable verification of results, and permit new innovative research built on existing information. Fundamentally, the two definitions of RDM consist of different activities and processes associated with data creation, storage, security, preservation, retrieval, sharing, and reuse taking into account technical capabilities, ethical considerations, legal issues, human resource capability and government frameworks.

Pinfield, Cox and Smith (2014) opined that RDM is a requirement in research institutes due to vast quantities of born-digital and non-digital research data which are now being produced in a wide variety of forms and at a rapid rate creating challenges of data called “data deluge”. This data deluge generates a need to develop policies, infrastructure, and services to manage data with the aim of assisting research institutes in creating, appraising, describing, preserving, accessing, reusing and sharing research data. RDM brings benefits to researchers and research institutes in many ways (Lewis, 2010; and Dora and Kumar, 2015) such as:

i. Ability to share research data, minimizing the need to repeat work in the field or laboratory;

ii. Research data gathered at considerable cost is not lost or inadvertently destroyed;
iii. Retrieval, comparison, and co-analysis of data from multiple sources can lead to powerful insights;
iv. New research themes can emerge from re-analysis of existing data or comparisons with new data;
v. Long-term preservation of data provides for validation check of the data and this enhances the credibility and transparency of the research data used;
vi. By opening research data sets for the public, there is visibility of the host institution and its researchers;
vii. Research funders are increasingly requiring researchers to deposit their research data for proper curation, full utilization, preservation, and reuse (Heidorn, 2011; Ingram, 2016; Lyon, Patel, and Takeda, 2014).

Furthermore, Ray (2014) explains that the sharing of research data increases the return on large investments, advances human knowledge, promotes economic development and reduces costly data duplication. Open access has emerged as one way of sharing research data to promote the advance of knowledge and technology transfer.

1.1.1 RDM Perspectives

RDM in developed countries such as United Kingdom (UK), United States of America (USA), Australia and Canada have made great advancements (Lewis, 2010; National Science Foundation (NSF), 2007a; and Henty, 2014). Large-scale management of the research data emerged over forty years ago in Europe when the UK Data Archives was established to manage paper-based surveys and other data outputs. This has been given impetus by the growth of digital research data and growing interest in long-term preservation, curation, and storage of research data for reuse and sharing (Lewis, 2010). In the year 2001, the UK government funded e-Science Core Program, administered by the Engineering and Physical Sciences Research Council (EPSRC) on behalf of Research Councils UK to establish infrastructure, middleware and documentation to facilitate wide uptake of RDM (Lewis, 2010; and Hey and Trefethen, 2003). Furthermore, e-Science Core Program also supported demonstrator projects to enable researchers to understand the scope, capability, and implications of e-research projects and the need to manage data that was generated forthwith with a focus on areas (Lewis, 2010) such as:

- Data-intensive: generating and often using large volumes of data;
- Collaborative: involving researchers across multiple institutions and transnational limitations;
- Grid-enabled: using high-capacity network and middleware.

According to Lord and Macdonald (2003), the Joint Information Systems Committee (JISC) commissioned a report on the curation of e-Science data and together with the e-Science Core Program, highlighted the role of the Digital curation centre with recommendations about the need to develop national capacity and capability to handle RDM. As can be seen, RDM featured prominently in e-Science Core Program and JISC in the UK which propelled the significance of Data Curation, (Data Curation Centre, DCC) lifecycle model (was a key recommendation in the JISC), technical infrastructure, legal issues, and human capabilities.

The growth of digital research has seen the emergence of data-intensive and collaborative research leading to the establishment of the National Science and Technology Council Committee (NSTCC) in the USA and the e-Infrastructure Reflection Group in the European Union to advise on capability, capacity and infrastructure in data management (Van den Eynden, Corti, Woollard, Bishop and Horton, 2011). These developments have increased investment in data management (Lewis, 2010). For example, the USA- NSF has invested funds and cyber-infrastructure for research data curation through DataNet programme (NSF, 2007a). The introduction of DataNet program was as a result of the introduction of data management plan requirements by the NSF which could result in an effective and efficient RDM and for this reason drew attention to the need for data management infrastructure, both in terms of hardware, human and policy support (Halbert, 2013; and NSF, 2007a). The NSF mandate was neither unprecedented nor an isolated intervention. The National Institutes of Health (NIH) had implemented the first mandate in 2003, requiring researchers to comply with data sharing and data management practices (NIH, 2003a). Other federal agencies, for example, the National Endowment for the Humanities, adopted a requirement for data management plan that explicitly emulated the NSF requirement (National Endowment for the Humanities, 2013).

The NSTCC on Science set up an Interagency Working Group (IWG) on digital data in 2007. The purpose of the IWG was to develop and promote the implementation of a strategic plan for the Federal government to cultivate an open interoperable framework to ensure reliable preservation and effective access to digital data for research, development, and education in science, technology and engineering (Interagency Working Group on Digital Data, 2009). The IWG did set out a roadmap for a series of coordinated national activities and included the clear statement:

> We envision a digital scientific data universe in which data creation, collection, documentation, analysis, preservation, and dissemination can be appropriate, *reliably, and readily managed. This will enhance the return on our nation's* research and development investment by ensuring that digital data realize their full potential as catalysts for progress in our global information society (Interagency Working Group on Digital Data, 2009).

To this end, it can be noted that the USA government has laid down strategies to create a comprehensive framework of transparent, evolvable and extensible policies, infrastructure, management and organizational structures that provide reliable and effective access, reuse and sharing of research data.

Australia has also moved with relative alacrity to develop data management of e-research and has set up the Australian National Data Services (ANDS) funded by the Australian government through the National Collaborative Research Infrastructure Strategy (NCRIS) (ANDS Technical Working Group, 2007). In this regard, ANDS core purpose is to make Australia's research data assets more valuable for researchers, research institutions and the nation. Henty (2014), in support of this view opines that Australian research institutions and researchers are keen to have greater access, sharing, and reuse of research data resulting in greater efficiency in RDM. In light of newly developed strategy, "The 2011 Strategic Roadmap for Australia Research Infrastructure", the Australian government has made significant investments in research data infrastructure to facilitate collection, generation, manipulation, curation, access and dissemination (Australian Government, n.d). Carrick (2014) points out that RDM in Australia is an essential component of all research leading to the establishment of, and sharing of 'Australian Code for the Responsible Conduct of Research', jointly developed and issued by the National Health and Medical Research Council (NHMRC), the Australian Research Council (ARC), and Universities Australia. The code assigns researchers and their institutions the responsibility of addressing ownership, storage and retention, access to, and sharing of research data.

In Canada, RDM services have become a high priority for government agencies and post-secondary institutions in recent years. While it has lacked coherent national strategies for developing the digital infrastructure required for e-research, such as those in Australia or the UK, still there remains a growing expectation for sound RDM (Whitehead and Bourne-Tyson, n.d). Due to different aspects of digital infrastructure being operated separately at different government levels and not as part of a cohesive whole at the national and institutional levels, the Canadian government funded three main federal research granting council known as the ‘Tri-Agencies’ with the sole purpose of strengthening RDM in Canada and maintain Canada’s research excellence (Government of Canada, 2016). The Tri-Agencies (Whitehead and Bourne-Tyson, n.d; and Government of Canada, 2016) are:

a) The Canadian Institutes of Health Research;
b) The Natural Sciences and Engineering Research Council of Canada; and
c) Social Sciences and Humanities Research Council (SSHRC).

According to Fry, Doiron, Létourneau, Perrier, Perry et al. (2017), the Tri-Agency Statement of Principles on Digital Data Management heightens the need for a collaborative national perspective on RDM that has been missing in Canada. Under its auspices, the statement outlines the agencies’ overarching expectations for RDM and the role of researchers, research institutions, research communities, and research funders in supporting data management (SSHRC, 2015). In Addition, the Canadian government through the Tri-Agencies promotes and supports research, research training, knowledge transfer and innovation within Canada. Like other developed countries, the Canadian government and SSHRC (Government of Canada, 2016; and SSHRC, 2015;) maintains that in promoting access to research results, the Tri-Agencies aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. In this respect, the Canadian government and research funders are becoming increasingly aware of the value of research data, the importance of fostering reuse of research data and the need for policies to enable excellence in RDM (Government of Canada, 2016; and Sewerin, 2015).

South Africa is leading the cluster of African countries in embracing RDM (Van Deventer and Piennar, 2015). Some research councils and institutes, and academic and research libraries in South Africa have initiated programs towards the realization of RDM. For example, the University of Cape Town (UCT) has established e-Research centre to work and partner with researchers in finding IT solution for their research work while the University of South Africa (UNISA) has completed investigation into RDM as part of the plan to establish data management (Macanda, Rammutloa and Bezuidenhout, 2015). The University of Pretoria, Stellenbosch and Witwatersrand are at different planning and implementation stages (Van Wyk and Van der Walt, 2014; Van Deventer and Piennar, 2015). An investigation on “Research Data Management in South Africa” by Kahn, Higgs, Davidson and Jones, (2014) found that in South Africa, a number of data repositories have been established to manage research data. They include South African National Park, National Health Information Repository and Data Warehouse, and Data Intensive Research Initiative of South Africa (DIRISA). Lötter (2014) and Fernihough (2011) affirm that DIRISA is one of the initiatives aimed at promoting RDM in the country.

Van Deventer and Piennar (2015) postulates that the Council for Scientific and Industrial Research (CSIR), DIRISA and National Integrated Cyberinfrastructure System (NICIS) are setting systems in place in terms of data curation, policies, technical infrastructure, and human capability to enable management, reuse, and sharing of research data. In the same vein, the National Research Foundation (NRF) - South Africa, as the leading government research funding agency, has been involved in many initiatives to allow the sharing of research outputs, datasets research support and knowledge networking databases which contribute to knowledge generation for the support and promotion of research development (NRF, 2015). Studies done by Kahn et al. (2014); Lötter (2014); and Van Deventer and Piennar (2015) on RDM landscape in South Africa, shows that there are high levels of awareness of RDM in research and academic institutions with notable activities of Network of Data and Information Curation Communities (NeDICC) in most research institutions. NeDICC aims to promote the development and use of research data including curation standards and practices of research institutions to ensure the long term preservation and accessibility of digital research outputs (UCT, 2016). Kahn et al. (2014) nevertheless singled out issues of leadership, policy and skills as areas that require the most urgent attention.

In Kenya, there is some attempt, albeit limited, to promote RDM especially in the health and migration sector (Jao, Kombe, Mwalukore, Bull, Parker et al., 2015; Family Health International-Kenya, 2005; Olum, 2013). Olum (2013) posits that despite Kenya having adequate migration data in various institutions, the data is not sufficiently coordinated, shared, analyzed or disseminated. Olum further observed that many of the government and civil institutions have limited capacity, resources and facilities for collection, analysis, use and reuse, and sharing of migration data hence making access and use of migration data difficult. Furthermore, the health sector in Kenya is making strides in embracing the sharing of research data albeit at a rate of limited rapidity. Lairumbi, Parker, Fitzpatrick and Mike (2011) opined that health research data sharing in Kenya is vital among health researchers for example, sharing research data on malaria or HIV/AIDS could enhance clinical vaccine trials and advance breakthroughs in the health sector. However, sharing health research data is not fully embraced, leading to underutilization of research data.

World Agroforestry Centre based in Kenya organized RDM training for agroforestry research scientists whose aim was to encourage necessary allocation of resources such as, skilled personnel and technical infrastructure for data management (World Agroforestry Centre, 2002). Equipped with the RDM training manual, the World Agroforestry Centre made an incredible step in introducing RDM into the agriculture sector. The assessment by scholars (Chisenga, 2012; Mugata, 2014; Alila and Atieno, 2006; World Agroforestry Centre, 2002) on RDM in Kenya established that various institutions, specifically agricultural research institutes have rich valuable research data that needs to be managed effectively in order to enhance the institutes mandate, increase agricultural productivity and enhance food security. Mugata (2014) asserts that accessing agricultural research data in Kenya is not easy for researchers and other stakeholders in agriculture due to limited institutional skills, legal framework, infrastructure and strategies that support RDM. However, to facilitate agricultural related content accessibility, visibility and sharing in Kenya, Kenya Agricultural Information Network (KAINet) was established (Mugata, 2014). KAINet aims at building a common and freely accessible information system for the generation, collection, processing, preservation and dissemination of agricultural research data and information.

This study focuses on RDM in Kenya's agricultural research institutes and Sandler (2013); and Alila and Atieno (2006) defines agriculture as the science or practice of farming, involves cultivating animals, plants, fungi, and other forms for food, fiber, biofuel, and other products used to sustain life. On the other hand, Herren and Donahue (1991); and Burton (2010) define agriculture as the broad industry engaged in the production of plants and animals for food and other resources, the provision of agricultural supplies and services, and the processing, marketing and distribution of agricultural product. Thus, agriculture could be referred to as the production, processing, promotion and distribution of agricultural products.

The Agricultural sector plays a strategic role in the process of an economic system of any given country; as a source of livelihood, contribution to national revenue, employment opportunities, food security, and foreign exchange resources. Kenya national economic growth is highly correlated to growth and development in agriculture, contributing 24 percent of national Gross Domestic Products (GDP) directly and another 27 percent indirectly (United Nations Environment Programme (UNEP), 2015). Moreover, Kenya Vision 2030 (Kenya, Republic of, 2007) identified agriculture as one of the key sectors to deliver a 10 per cent annual economic growth rate, therefore, National Agricultural Research Systems (NARS) should ensure that information generated by agricultural research institutes is collated and made available to the agricultural stakeholders (Kenya, Republic of, 2012). Manda (2002) emphasizes that unless agricultural research data is managed well and extension institutions are transformed, information will play a marginal role in the process of agricultural transformation.

Agricultural research, on the other hand, can be defined as an activity aimed at improving productivity and quality of crops and animals through their genetic improvement, better plant protection, irrigation, storage methods, farm mechanization, efficient marketing and better management of resources (Loebenstein and Thottappilly, 2007). Kenya, Republic of ?(2012) points out that Kenya agricultural research is geared at enhancing productivity, product quality, and safety as well as competitiveness in domestic and global markets. In this regard, NARS policy was established to reform the Kenya agricultural research systems into a dynamic, innovative, responsive and well-coordinated system driven by a common vision and goal (Kenya, Republic of, 2012) which resonated with efficient and effective RDM.

The continuing explosion of agricultural research data within the global agricultural research landscape and the increase interest in sharing and curating research data has triggered the emergence of RDM as a service area supporting researchers in different organization in Kenya (Agricultural Information Management Standards (AIMS), 2017). AIMS is earmarked for active progression in alignment and sharing of trusted, findable, accessible, interoperable and reusable research data in all agricultural research domains (AIMS, 2017).

1.2 Kenya Agricultural and Livestock Research Organization (KALRO)

The focus of this study on agriculture is premised on the fact that Kenya’s, agricultural sector is the mainstay of the country’s economy because it contributes 26% of the GDP and accounts for 65% of the country’s total export and provides more than 18% of formal and 70% of informal employment respectively in the rural areas (Kenya, Republic of: Ministry of Agricultural, Livestock and Fisheries, 2010; UNEP, 2015; Kenya Agricultural Research Institute (KARI), 2012). The Kenya government therefore attaches great importance to the agricultural sector, consequently there is a need to invest in agricultural research to boost productivity, economic growth and food security.

Kenya agricultural research has undergone tremendous changes since its inception early in the 19th century by the colonial government, with formal agricultural research in Kenya dating back to 1903, the establishment of an agricultural testing station at Kabete and livestock station in Naivasha (Kenya, Republic of, 2012). After the breakup of the East Africa Community 1975, the Science and Technology Act Cap 250 was passed in 1979 as a basis for establishing the National Council for Science and Technology and a number of institutes to carry out research in the country (Kenya, Republic of, 2012). Since then Kenya agricultural research has grown tremendously leading to the establishment of the Kenya KARI through an Act of Parliament (Cap 250, revised 1979) which became operational in 1986 with an aim of improving agricultural productivity and subsequently contributing to improved food production (Miruka, Okello, Kirigua and Murithi, 2015). In the meantime, KARI became one of the most developed systems in sub-Saharan Africa in terms of human and physical capacity as well as the pool of knowledge and technology in agricultural research (Kenya, Republic of, 2012).

In implementing the Agricultural Sector Development Strategy 2010-2020, the Kenya government reformed the NARS(s) through the creation of Kenya Agricultural and Livestock Research Organization (KALRO). KALRO is a corporate body and was set up vide Kenya Gazette Supplement No.29 (Acts No. 17): The Kenya Agricultural and Livestock Research Act, 2013 No. 17 of 2013, with appointment of KALRO Board to coordinate agricultural research in the country (KALRO, 2016; Kenya, Republic of, 2012). The Act however, does not clearly define how research data generated in the research institutes should be managed to ensure the continued preservation, long-term access, sharing and reuse of the data. KALRO was launched on 18 August 2014. Its formation was aimed at restructuring agricultural and livestock research into a dynamic, innovative, responsive and well-coordinated system driven by the following KALRO, 2016 strategic objectives (KALRO, 2016)**:**

a) To generate and promote technologies and innovations for demand-driven agricultural and livestock product value chains;

b) To develop and promote markets and marketing strategies for agricultural and livestock product value chains;
c) To undertake policy analysis and advocate policy option for enhancing demand-driven agricultural and livestock product value chains;
d) To strengthen the capacity for implementing agricultural and livestock products value chains;
e) To enhance availability of knowledge, information, and technologies on agricultural and livestock product value chain;
f) To enhance good corporate governance.

Core functions

The core functions of KALRO are (KALRO, 2016):

a) Formulate policy and make policy recommendations to the Cabinet Secretary on agricultural research;
b) Prioritize areas for, and co-ordinate, agricultural research in Kenya in line with the national policy on agricultural;
c) Determine and advise the Government on the resource requirements for agricultural research in Kenya both at national and county level;
d) Regulate, monitor and ensure that all agricultural research undertaken by research institutes and other institutions or persons undertaking agricultural research is consistent with the national priorities specified in the relevant policy documents;
e) Establish and excise control over the research institutes, committees and research centers established pursuant to this Act;
f) Formulate or approve medium and long term research plans, strategies and budgets of research institutes, committees and organization's established pursuant to this Act;
g) Provide grants to research institutes and persons desirous of carrying out research and training programs which are consistent with the national research priorities and plans of the organization;
h) Support and promote the training and capacity building in relation to agricultural research;
i) Promote the dissemination and application of research findings in the field of agriculture and the establishment of a Science Park;

j) Liaise with and ensure the co-ordination of institutions, agencies and persons involved in agricultural research;
k) Establish platforms for the purposes of sharing research information, advancing research and transfer of technology and dissemination of information relating to advancements made in agricultural research;
l) Ensure continuance of performance improvement in the field of agricultural research; and
m) Perform such other functions as may be conferred on it by this Act or any other written law.

KALRO is the premier national institution that brings together research programs in food crops, horticultural and industrial crops, livestock and range management, land and water management and social-economics. Additionally, KALRO promotes sound agricultural research, technology generation and dissemination to ensure food security through improved productivity and environment conservation (Devex, 2017).

The human resources at KALRO is composed of staff who are competent, qualified, skilled, distinguished scholars and researchers who work tirelessly towards the realization of the organizations, vision, mission and core values (KALRO, 2016). The research institutes under KALRO have resources and services which are open to stakeholders for research and academic visits for the betterment of the knowledge sharing (KALRO, 2016; and Devex, 2017).

KALRO has currently sixteen (16) agricultural research institutes (KALRO, 2016) namely:

1. Food Crops Research Institute
2. Coffee Research Institute
3. Biotechnology Research Institute
4. Genetic Resource Research Institute
5. Apiculture Research Institute
6. Arid and Range Lands Research Institute
7. Beef Research Institute
8. Dairy Research Institute
9. Horticulture Research Institute
10. Industrial Crops Research Institute
11. Sugar Research Institute
12. Tea Research Institute
13. Sheep and Goat Research Institute
14. Veterinary Research Institute
15. Agricultural Mechanization Research Institute
16. Non-Ruminant Research Institute

The functions of the research institutes under KALRO (KALRO, 2016) are to:

a) Advise on, and develop appropriate systems to promote balanced, diversified and sustained agricultural development and to optimize agricultural production through adaptive and investigative research; and
b) Facilitate the use of improved production technology, and to establish adequate feedback systems from agricultural producers in order to achieve and maintain national self-sufficient and export capacities in agricultural products.

Therefore, it can be noted that RDM is imperative in agricultural research institutes to promote access, sharing and reuse.

1.3 Statement of the problem

Agriculture is the bastion of Kenya's economy and a major contributor to national food security and a stimulant to the growth of employment. The Kenya government strategy for revitalizing agriculture links the national research system with the agriculture sector (Kenya, Republic of: Ministry of Agricultural, Livestock and Fisheries, 2010; Kenya, Republic of, National development Plan, 2002-2008). KALRO was therefore established through an Act of parliament (The Kenya Agricultural and Livestock Research Act No.17 of 2013) to coordinate agricultural research in the country (KALRO, 2016). The Kenya Agricultural and Livestock Research Act No.17 of 2013 however, do not define how research data generated by the research institutes should be managed to ensure the continued preservation, long-term access, sharing and reuse of such research data. Furthermore, the two functions of the research institutes under KALRO as stipulated by the KALRO Service charter (KALRO, 2016) does not mention RDM making the full implementation or practice of RDM a challenge.

In addition, the agricultural research institutes in Kenya generate a lot of research data however little is known about the mechanisms for the management of such data especially with regard to curation, sharing and reuse (Alila and Atieno, 2006). Mugata (2014) asserts that though the challenges of RDM in the agriculture research institutes in Kenya are known, it remains unclear as to why they have not been addressed, resulting in poor mechanisms for data curation, sharing and exchange, low quality of research outputs, duplication of research, high costs of gathering data, and poor re-analysis of existing research data. Added to which, a framework for capturing, organizing, and preserving data for the long term is nonexistent resulting in valuable datasets becoming lost or discarded when researchers leave or disengage with the research institutes as mechanisms for managing succession are dysfunctional. The situation is not made any easier as RDM legal, policies and guidelines to encourage researchers to deposit their research output in appropriate spaces such as the institutional repositories are non-existent or not enforced where they exist (Beintema, 2015; Alila and Atieno, 2006). Halbert (2013) underscores the importance of researchers sharing cumulative sets of research data with other multi-institutions to advance larger research agenda for the wider public good.

Additionally, Wambani (2011) in a study found that researchers in Kenya were inadequately supported and trained to effectively improve data collection, appraisal, preservation, access, sharing and reuse. These circumstances have resulted in incomplete and inaccurate data, along with loss of research data consequently hampering access, sharing, use and reuse of research data (Ndemo, 2016). Kenya, Republic of, (2012) further, indicated that despite the large number of Kenyan skilled scientific staff engaged in agricultural research in both public and private institution, no mechanism exists to harness these strengths at national level. It is argued that although Kenya's agricultural research institutes carry out research, little is known about how such data is curated, shared and reused. The World Agroforestry Centre (2012) corroborates this claim saying little has been documented on RDM in Kenya. This claim is further enhanced when other agricultural research institutions (universities and International/regional organizations) carry out agricultural research independently without a strong and clear identifiable coordinating body. The challenge is how to establish an integrated RDM system that is well balanced and directed to address the diversity of agricultural research data in Kenya in light of limited resources (Kenya, Republic of, 2012).

Little research seems to exist on the subject of RDM in the agricultural research institutes in Kenya as revealed by a search on databases such as Centre for Agriculture and Biosciences International (CABI), Agricola, Agriculture Journals and Food and Agriculture Organization (FAO). This study therefore addresses the major research question: How is research data managed in Kenya's agricultural research institutes to promote access, sharing, reuse and disposal of research data?

1.4. Objectives of the study

The study addresses the following two broad objectives:

a) To assess the status of research data management in Kenya's agricultural research institutes;
b) To determine the legal and policy framework, Information and Communication Technology(ies) (ICTs) infrastructure and human capital that is available or not to facilitate RDM in Kenya's agricultural research institutes.

1.4.1 Research questions

In order to address the objectives of the study, the following research questions were addressed:

1. How does the availability or absence of legal, policy and regulations affect the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes?
2. How do Kenya's agricultural research institutes capture, appraise, describe, preserve, and make accessible for reuse its research data?
3. What knowledge, skills and training are needed to capture, appraise, describe, preserve, and make accessible for reuse its research data?
4. What is the level of ICT preparedness in Kenya's agricultural research institutes for the capture, appraisal, description, preservation, access and reuse of research data?
5. How do collaborative partnerships influence the capture, appraisal, description, preservation, access and reuse of research data in Kenya's agricultural research institutes?

1.5 Significance of the study

A review of the literature brings to the fore the significance of the study to typically include an explanation of the work's significance, its potential benefits and its overall impact which attempts to explain to the audience why a researcher's work is worth performing and how the study will fill in knowledge gaps in their field (Creswell, 2014; Bryman, Teevan and Bell, 2009). Although the agricultural sector in Kenya is the foundation of national economy, agricultural research division has not been fully exploited in terms of developing or improving products and technologies which add value to existing ones and are of significant importance to the end user. The current study was necessitated by the fact that RDM in Kenya's agricultural research institutes is the source of knowledge and innovations that drive current and future agricultural development. For this reason, there is a need for RDM to facilitate research data capture, description, preservation, accessing, use and re-using, and sharing. In essence, sound RDM can benefit research institutes and assist researchers in avoiding duplication. Additionally it can lead to powerful insights through retrieval, comparison and co-analysis of data, emergence of new research themes from re-analysis of existing data and avoid lost research data gathered at considerable cost.

This study contributes to the existing body of knowledge by consolidating RDM capability factors such as skills and training, technical infrastructure, legal and policy frameworks, and collaborative partnership and data curation (capture, describing, preservation, access, and use and reuse). Moreover, extant literature revealed that RDM in agricultural sector and especially in Africa is a relatively new area of study (The World Agroforestry Centre, 2012; Kahn et al, 2014; Mugata, 2014, and Wambani, 2011) and therefore this study will make a contribution to the body of knowledge in this area from the context of a developing country.

This study is expected to inform and expand knowledge-base associated with RDM legal and policy framework by creating awareness among RDM stakeholders about the need to align RDM legal and policy framework with agricultural research institutes. This alignment is essential in guiding the stakeholders on agricultural research data capture, appraisal, description, preservation, access, use and reuse, and sharing. Additionally, the study findings will inform the formulation of data curation policy and policies on human resource capabilities, technical infrastructure, and collaborative partnership among others. The findings of the study will be useful to government, funding agencies, directors of agricultural research institutes, researchers, librarians and IT specialists in their current and potential roles in RDM.

1.6 Delimitations of the study

The focus of this study; 'RDM on Kenya's agricultural research institutes', sought to gain an insight on how agricultural research data is generated, appraised, described, preserved, accessed, used and reused, and shared. The other focus of the study was on RDM legal and policy framework, human resource capability, technical infrastructure and collaborative partnerships. This brought to the vanguard RDM in Kenya agricultural research institutes.

The study purposely target six (6) Kenya's agricultural research institutes under KALRO namely; Food Crops Research institute (FCRI), Coffee Research Institute, Tea Research Institute (TRI), Dairy Research Institute (DRI), Biotechnology Research Institute (BRI), and Genetic Resource Research Institute (GeRRI) (KALRO, 2016). The selected agricultural research institutes are situated in areas where agricultural activity is intensive and have a long history of undertaking agricultural research in Kenya in different agricultural spheres of livestock, food crops, cash crops, biotechnology and more.

The population of study comprised the agricultural research institute's director, heads of research, researchers, librarians and head IT department. The study was limited by the tight schedules among some of the 'would–be' participants of the study, especially the directors of the institutes and the researchers. However, a vigorous follow up helped the researcher obtain a critical mass of data for the study. Moreover, due to financial constraints the researcher selected only six (6) agricultural research institutes.

1.7 Research methodology

A detailed Research Methods is covered in chapter four discussing issues of methods and research design. Facets of methodology are covered here just to introduce key subjects that are discussed in chapter four. The study adopted pragmatism ontology. Pragmatism focuses attention on the research problem and the use of pluralistic approaches to derive knowledge about the problem (Rossman & Wilson, 1985). The study applied mixed methods epistemology that enabled the researcher to collect quantitative data from a large sample of researchers from Kenya's agricultural research institutes and also qualitative data from directors of institutes, heads of research, heads of IT, librarians and archivists. The use of mixed methods epistemology ensured that data collected through one method could be validated using the other method (Creswell & Plano Clark, 2011).

The study purposively targeted five (5) Kenya's agricultural research institutes and the population was stratified into six i.e. directors of institutes, heads of research, researchers, librarians and heads of IT.

Validity of the research instruments was achieved through face and content validity by linking the questions and the objectives of the study (Kumar, 2011). Reliability of the questionnaire was achieved through pilot study, Cronbach values of above 0.7, linking the questions and the objectives of the study and triangulation of data sources.

The study complied with UKZN research ethical guidelines. In addition, a research permit was sought from National Commission for Science, Technology and Innovation (NACOSTI) in Kenya. Further permission was sought from the KALRO where the study was undertaken.

1.8 Structure of the study

The study is divided into seven chapters as follows:

Chapter one: Background to the Study

This chapter covers background to the study, research problem, research objectives, research questions, significance of the study, and delimitations of the study.

Chapter Two: Theoretical Framework

This chapter presents detailed description of theoretical framework and underpins the study.

Chapter Three: Literature Review

This chapter provides a detailed review of both theoretical and empirical literature in books, journal articles, online databases and other sources. The gaps in the literature are identified and the contribution of this study in addressing them adduced.

Chapter Four: Research Methodology

This chapter presents in detail: research paradigm, research approach, research design, study population, sampling technique, sample size, data collection methods, data analysis, validity and reliability of the data collection instruments, and ethical considerations.

Chapter Five: Data Analysis and Presentation of the Findings

This chapter presents findings of the study, guided by theory and research questions.

Chapter Six: Discussion of Findings

This chapter interprets and discusses the findings of the study using the theoretical lens that underpinned the study and also extant literature. The originality and contribution of the study is provided.

Chapter Seven: Summary, Conclusion and Recommendations

This chapter provides a summary of findings, conclusion and recommendations. Further areas for research are presented.

Summary

The chapter introduced and presented the conceptual setting of the study. The chapter defined agricultural research data, RDM and by extension discussed its role in Kenya agricultural research institutes. The contextual setting of RDM was discussed and Kenya’s agricultural research institutes were the focus point to determine the role and RDM in Kenya’s agricultural sector. The chapter also provides the statement of the problem, motivation of the study, purpose of the study, research objectives, and research questions. Further, the chapter provides significance and delimitations of the study, a brief introduction of theoretical framework, literature review and research methodology adopted in the study. The proposed structure of the thesis was also outlined. The next chapter provides detained discussion of the theoretical framework if the study.

CHAPTER TWO

THEORETICAL FRAMEWORK

2.1 Introduction

Theory guides every aspects of research, from formulation of the research question to operationalization and discussion. Theories are formulated to explain, predict and understand phenomena and in many cases to challenge and extend existing knowledge within the limits of critical bounding assumption (Mason and Culnan, 1995; Bernath and Vidal, 2007). According to Kerlinger (1979), a theory is a set of interrelated constructs (variables), definitions and propositions that presents a systematic view of phenomenon by specifying relations among variables, with the purpose of explaining natural phenomena. Garrison (2000) on the other hand, defined a theory as a coherent and systematic ordering of ideas, concepts and models, with the purpose of constructing meaning to explain, interpret and shape practice. In this respect, a theory not only explains known facts, it also allows researchers to make predictions of what they should observe if a theory is true.

According to King, Keohane, and Verba (1994), there is no empirical investigation that can be successful without theory to guide its choice of question. Therefore, the purpose of a theory or theories in this study as documented by May, 1993; Neuman, 2000 and Creswell, 2009 is to:

a) Give the inquiry a focus;
b) Prevent the fragmentation of knowledge by ordering;
c) Provide theoretical explanations and deeper understanding of what is being investigated;
d) Provide tools for the interpretation of collected data.

Theoretical framework provides scientific justification for investigation by showing that the research does not appear suddenly from “out of the blue”; rather it is grounded in and based on scientific theory (Swanson, 2013). Furthermore, the theoretical framework introduces and describes the theory that explains why the research problem in this study exists and indicates theories and analytic models that are relevant to the research problem being investigated. Sekaran (2003) defines a theoretical framework as a conceptual model of how one theorizes or makes logical sense of the relationships among several factors that have been identified as important to a problem. The purpose of a theoretical framework is to make research findings meaningful and generalisable thus stimulating research and the extension of knowledge by providing both direction and impetus (Polit and Beck, 2004).

The University of Southern California (2016) outlines the following motivations for using theoretical framework:

1. Connects the researcher to existing knowledge. Guided by a relevant theory, the researcher is given a basis for hypothesis and choice of research methods;
2. Helps the researcher identify the limits to those generalizations. A theoretical framework specifies which key variables influence a phenomenon of interest and highlights the need to examine how those key variables might differ and under what circumstances;
3. Articulate the theoretical assumptions of a research study by forcing the researcher to address questions of why and how;
4. Provides an explicit statement of theoretical assumptions to permit the researcher to evaluate them critically.

Theoretical framework should fit its purpose in order for it to effectively inform an inquiry. Therefore, the development of a theoretical framework to guide researcher's research is the central piece in the research puzzle this explains why Ennis (1999; and Maxwell, 2013) suggests that the most critical part of the research plan is the theoretical framework.

2.2 Theories underpinning the study

The purpose of this study is to examine Research Data Management (RDM) practices in Kenya's agricultural research institutes with the view to propose interventions to improve management, sharing and reuse of agricultural research output.

The following research questions were addressed:

1. How does the availability or absence of legal, policy and regulations affect the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes?
2. How do Kenya's agricultural research institutes capture, appraise, describe, preserve, and make accessible for reuse its research data?
3. What knowledge, skills and training are needed to capture, appraise, describe, preserve, and make accessible for reuse its research data?

4. What is the level of ICT preparedness in Kenya's agricultural research institutes for the capture, appraisal, description, preservation, access and reuse of research data?
5. How do collaborative partnerships influence the capture, appraisal, description, preservation, access and reuse of research data in Kenya's agricultural research institutes?

There are various theories that are relevant for investigating research data management that include among others, Community Capability Model (CCM) Framework, Data Curation Centre (DCC) Lifecycle Model, A library-Oriented Model of Institutional RDM and The Open Archival Information System (OAIS) Functional Model.

The study was primarily underpinned by the Community Capability Model (CCM) Framework and Data Curation Centre (DCC) Lifecycle Model.

2.2.1 Community Capability Model (CCM) Framework

Community Capability Model (CCM) framework was developed by United Kingdom Office for Library and Information Networking (UKOLN), University of Bath and Microsoft Research to assist research funders, institutions and researchers in growing the capability of their communities to perform data-intensive research (Lyon et al., 2012) by:

a) Profiling the current readiness or capability of the community;
b) Indicating priority areas for change and investment; and
c) Developing roadmaps for achieving a target state of readiness (Lyon et al., 2012).

The CCM framework comprises eight capability factors that include collaboration, skills and training, openness, technical infrastructure, common practices, economic and business, legal and ethical, academic issues representing human, technical and environmental issues.

The capabilities in CCM framework are very crucial in RDM especially in data-intensive disciplines like agriculture. Agricultural research data is generated in large scale and it has been given more impetus by the growth of digital research and the growing interest in long-term preservation, access for use and reuse. Managing agricultural research data is important to promote access, sharing and reuse. In order for this to happen, there is need to utilize the community/institutional capabilities, such as; skills and training, ICT infrastructures, legal and policy consideration, and collaborative partnership (Jones, Pryor and Whyte, 2013)

A study done by Bigagli, Sveinsdottir, Wessel, Smallwood, Linde et al., (2013) on policy recommendations for open access to research data revealed that vast amount of data are produced each day, which are neither discoverable, accessible nor reusable due to lack of curation, storage and overall management however, the role of infrastructure in RDM practices is seen to provide uniform and equal access to research data outputs to facilitated use, reuse and sharing. It is important in today's research environment especially in agricultural research institutes for researchers to have the ability to collect, analyze, share, and effectively manage and preserve agricultural research data. In contrast, capabilities related to supporting researchers in RDM have in many cases been found to be lacking (Newton, Miller and Bracke, 2011).

A well designed and coordinated community/institutional capability enhances and complements open access, sharing, use and reuse of research data at institutional and international level, derive collaborations among agricultural research institutions, government agencies, agricultural related industry and research community (Australian Government, n.d). Thanos (2010) opined that well-connected and positioned community/institutional capabilities such as human capital, legal and policy considerations, ICT infrastructure, and collaboration, will allow integration throughout the data lifecycle from collection, to processing, to preservation, access, use, reuse and sharing.

Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data (2016) observed that community/ institutional capabilities for RDM should include organizational practices, technical infrastructure and social forms that collectively provide for the smooth operation in RDM.

Community Capability Model (CCM) framework was developed to facilitate and attain an understanding of factors influencing RDM. Figure 2.1 below presents the CCM framework.

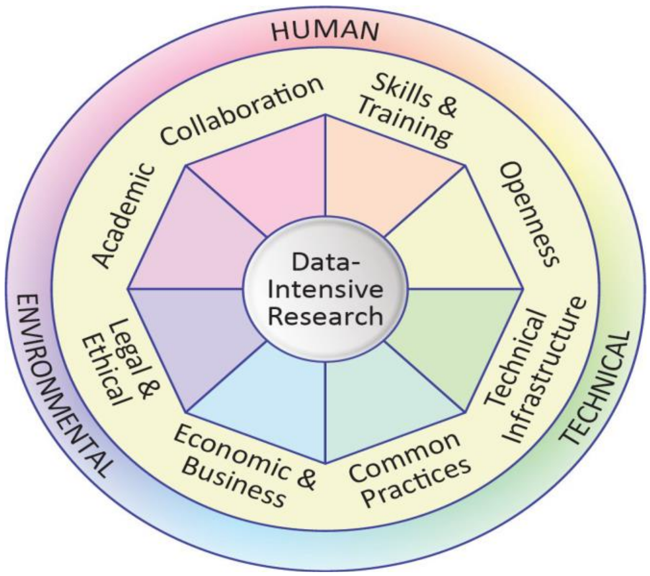

Figure 2.1: Community Capability Model (CCM) Framework (Source: Lyon et al., 2012)

The CCM framework is used to underpin the four research questions of this study and it has been found to be robust and successfully used in related studies of Crowston and Qin, (2012); and Lyon, Patel and Takeda (2014).

2.2.1.1 Key variables in CCM framework

(a) Openness variable

Open access to research data is critical for advancing science, scholarship and society. Arzberge, Schroeder, Beaulieu, Bowker, Casey et al., (2014) assert that open access to, and sharing of data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, facilitates the education of new researchers and permits the creation of new datasets when data from multiple sources are combined.

A study done by Antelman (2004) on the impact of open access revealed that open access articles have a greater research impact than articles that are not freely available and for this reason calls for initiatives such as building institutional repositories to facilitate sharing of research data. Currently research data are increasingly created in digital format and authors are encouraged to deposit the data that underpin their final articles in institutional repositories. On the other hand research funders require researchers to make data open access to allow sharing, use and reuse of research data to generate new data and knowledge (Enabling Open Scholarship, n.d).

Open access movement (Berlin Declaration, 2003; European Commission, 2016) have advocated for research data to be carefully preserved and made widely available through open access to enhance use and reuse of research data. Added to which, The Coherence in Information for Agricultural Research for Development (CIARD) (2012) asserts that research data output should be shared through the use of web 2.0 and social networking/media to improve both research data visibility and reuse. Denton Declaration (2013) stated that open access to research should be a central goal of the lifecycle approach to RDM and there should be infrastructural support for data discoverability, accessibility, share-ability, reuse and long-term stewardship. Denton Declaration (2013) posits that the principle of open access should not be in conflict with the intellectual property right of the researchers, instead a culture of citation and acknowledgement should be cultivated rigorously and conscientiously among all practitioners. Nonetheless, Bigagli et al., (2013) point out concerns that affect open access if infrastructure support is not implemented correctly to include poor heterogeneity and interoperability, accessibility and discoverability, preservation and curation, quality and security of research data.

(b) Skill and training variable

According to Henty (2014), the technical RDM skills include: high performance computing, information engineering, information modelling, portal design, database integration, metadata, and programming are vital while non-technical skills are: in information seeking, business analysis, project management communication and negotiation. Higgins (2012) outlines nine RDM workflow bearing roles, responsibilities and training requirements to include:

a) Data creation;
b) Assignment of persistent identifiers;
c) Metadata creation;
d) Attaching or linking the metadata to the data;
e) Indexing or tagging;
f) Quality assurance procedures for data, metadata and indexing;
g) Ensuring legal and ethical metadata is collected;
h) Appraisal and disposal procedures and
i) Procedures for transfer to storage media

Qin (2013) asserts that the personnel involved in RDM must be highly trained in technology, subject field and able to effectively communicate between different stakeholders. Schmidt and Shearer (2016) categorize the core competencies of RDM into three:

1. Providing access to data
   (i) Knowledge on: repositories, data discovery mechanisms, data manipulation, analysis techniques;
   (ii) Skill on: data organization, data licensing, intellectual property.
2. Advocacy and support for managing data
   (i) Knowledge on: funders' policies and requirements, data management plans, data publication requirements, data citation and referencing practices, best practices for data format, types and metadata;
   (ii) Skills on: articulating benefits of data sharing and reuse, data audit, assessment tools.
3. Managing data collections
   (i) Knowledge on: metadata standards and schemas, database design types and structure, data repositories and storage platforms;
   (ii) Skill on: select and appraise datasets, undertake digital preservation activities, activity manage research data.

RDM encompasses a wide array of activities across the research data lifecycle hence there is a need for RDM stakeholders to identify the required skill, knowledge and training for RDM staff and support units such as librarians, IT specialists and researchers.

(c) Technical infrastructure variables

The technical infrastructure is concerned with data capture, organizing, preservation, discovery and access, integration and collaboration platforms. Technical infrastructure is necessary to derive maximum benefits from data access and sharing. Arzberge et al., (2014) in this regard point out that technical infrastructure must be robust in terms of long term and diverse use, flexible to respond to the continuous and rapid changes in RDM and facilitate effective data access and sharing.

The technical infrastructure covers a wide range of technologies for collecting, storing, processing, organizing, transmitting, and preserving data as well as platforms for communication and collaboration (Qin, 2013; Smith, 2014). Included in this dimension of the research data infrastructure are networks, databases, web portals, repositories, web 2.0, social networks, authentication systems, research data management systems and software applications. Witt (2008) acknowledges that one piece of technical infrastructure that supports RDM is a distributed institutional repository that includes electronic documentations, digitized archival collections, and research datasets housed in multiple systems that are connected together using web services and other middleware. Moreover, research institution should have new mandates for RDM by means of increasing the demand for ICT infrastructure that can support RDM. The role of technical infrastructure is seen to be the provision of uniform and equal access to a broad variety of research outputs by making data understandable, searchable, retrievable, available, accessible, sharable and secure (Bigagli et al., 2013).

(d) Legal and policy issues

The legal and ethical issues variable is critical and is concerned with legal, policy and regulatory frameworks and ethical issues. RDM legal and policy frameworks are needed to guide the practices of research data creation, appraisal, description, preservation, access, sharing, reuse intellectual property, ethical issues, disposal among others (Anderson, 2004).

RDM legal, policy and relations are to guide understanding data types and formats that should be archived, policy on the use of metadata, whether the institution has a data repository for storing data files and policies covering its use, and what procedures should be established when sharing data (Qin, 2013; Mullins, 2014; Smith, 2014). Bohémier, Atwood, Kuehn and Qin (2011) identified six aspects of research data policies that should be addressed: data curation, management, use, access, publishing and sharing. Mossink, Bijsterbosch and Nortier (2013). MacKenzie (2014) on the other hand found that, while most funders of research projects will have policies covering data management, there is inadequacy of data management policies covering training and support, open access and reuse, security, data curation and preservation. Research data policies at national and institutional levels provide a framework for day to day operations of RDM with different policies addressing different areas of questions.

(e) Collaborative partnerships

The collaboration variable on the other hand caters for partnerships within the discipline/sector, across disciplines/sectors, and with the public. The University of Sheffield (2016) noted that collaboration is needed to facilitate sharing and reuse of research data in future research, strengthening of the research environment, improving research data workflows and visibility of research outputs.

More importantly, agricultural research institutes should encourage collaborative partnership on RDM at varying level ranging from international collaborations to national institutions, support units (Library, archives and (IT) department) and within individual research groups (Flores, Brodeur, Daniels, Nicholls and Turnator, 2015). National Library of Australia (2003) acknowledges that collaborative efforts can be highly advantageous to effective RDM, enabling the development and implementation of shared policy environments and documentation, workflows, tools and technical framework, staffing and costs.

2.2.1.2 Relevancy of the model to the study

CCM framework assists research funders, institutions and researchers in growing the capability of their communities to perform data-intensive research (Lyon et al., 2012). This implies that the usefulness of institutional capabilities relates to different stages of data lifecycle which is crucial in RDM. Cox and Pinfield (2014) observe that openness, skill and training, technical infrastructure, legal and policy issues and collaboration play a fundamental role capturing, appraisal, description, preservation, access and reuse of research data. CCM framework provides details of the roles, responsibilities and requirements of each capability for enhanced effective and efficient RDM.

The CCM framework focuses more on the adoption of ICT in every institutional capability due to the growth of digital research which is data-intensive in agriculture. This is particularly significant in generating vast agricultural research data, given that data curation deals with digital research data. Crowston and Qin, (2012) used the model to investigate data management assessment and planning tools in research data management. Lyon, Patel, and Takeda (2014) used the model to investigate requirements for research data management support in academic libraries, introducing a new multi-faceted capability tool.

2.2.1.3 Gaps in CCM model

CCM framework articulates well on developing institutional capabilities in addressing RDM; nevertheless the model does not focus in detail on data curation which is a significant constituent in RDM. Data Curation Centre (DCC) lifecycle model will therefore be used to complement CCM framework. DDC focuses on data curation (capture, appraisal, description, preservation, access and reuse).

2.2.2 Data Curation Centre (DCC) Lifecycle Model

The DCC was launched on 1 March 2004, following a successful response to Joint Information Systems Committee (JISC) Circular 6/038 by a consortium comprising the University of Edinburgh, University of Glasgow, UKOLN at the university of Bath, and Science and Technology Facilities Council (STFC). The development of DCC involved groups engaged in digital preservation and curation activities which included UK higher and further education, data specialists, records managers, librarians, archivists, researcher (as data creators), and policy makers (Data Curation Centre, 2004a). During its development the public and commercial sectors, international organization and standards working groups were also engaged.

The DCC lifecycle model promotes a lifecycle approach to the management of digital materials to enable their successful curation and preservation from their initial conceptualization to either disposal or selection for reuse and long-term preservation (Higgins, 2008). Data curation is about maintaining and adding value to a trusted body and appraisal of digital information over its entire life (Pennock, 2007). Figure 2.2 presents Data Curation Centre (DCC) lifecycle model.

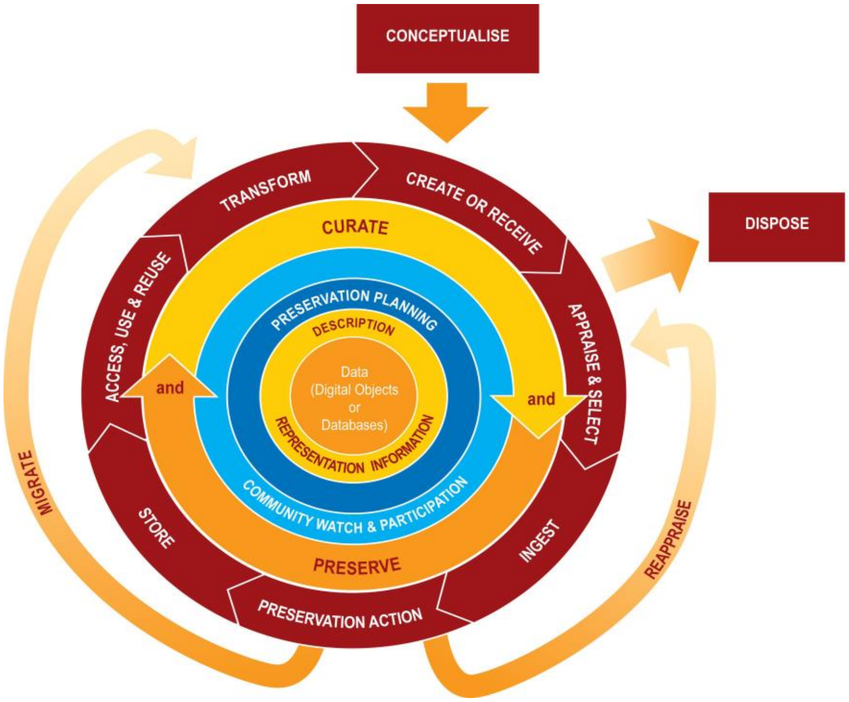

Figure 2.2: Data Curation Centre (DCC) Lifecycle Model (Source: Higgins, 2008)

Data curation is defined as the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose and available for discovery and reuse (Hinnant, Stvilia, Wu, Worrall and Burnett, 2012; Lord, MacDonald, Lyon and Giaretta, 2004; Laughton and Du Plessis, 2013). In the same vein, Yakel (2007:335) defines data curation as the active involvement of information professional in the management, including the preservation, of digital data for future use. Higgins (2011) asserts that technical development and mature understanding of data practices and procedures is necessary for ensuring access, use and reuse of digital data during its lifecycle.

Studies done by Tenopir, Birch and Allard (2012) on research data services revealed that researchers, librarians, IT specialists and archivists need to be trained in the area of data curation and management services for example, identifying and collecting data and datasets to include repositories, use of metadata, preservation, appraisal and selection. Developing an effective data curation plan requires agricultural domain-specific researchers, librarians, archivist and IT specialists to have an understanding of data curation requirements, practices and procedures. Further, Haas and Murphy (2009) emphasize that it is necessary for RDM stakeholders to get to know the data options and obligations of the disciplines they serve and get ready to facilitate communication before partnering with existing data sites, other research institutes at national and international level, as well as publishers. Data curation consists of a range of activities and processes focused on maintaining, preserving, and adding value to agricultural research data throughout its lifecycle which results in an effective way for addressing RDM (Pennock, 2007).

2.2.2.1 Key variables considered in DCC lifecycle model

According to Data Curation Centre (DCC) lifecycle model, data curation includes the data capture, appraisal, description, preservation, access, reuse and transformation of research data.

a) ***Data capture***

Data Capture involves how data will be captured and stored using descriptive and technical metadata. Building a collecting/capturing policy in order to prepare the receipt of research data from data creators is necessary. Higgins (2012) asserts that the data capture stage is critical because it is concerned with identifying how data will be created and by whom, and the documentation of this will form the basis of data curation lifecycle. The technologies and standards that are to be used should be equally identified and developed to make data capture and storage possible.

b) Data appraisal

Appraisal and selection is the process of evaluating research data in order to decide which to retain over the long term, which to retain for the meantime and which to discard according to appraisal and selection policies (Higgins, 2012). Whyte and Wilson (2010) outline five main reasons for undertaking the appraisal of research data namely to:

i. Reduce the amount of research data that has to be managed or curated over the long term with resources directed towards research data which has long term value;
ii. Facilitate the ability to maintain intellectual access to research data, to create and attach metadata, index and to store logically for data to be searched and retrieved quickly and efficiently;
iii. Ensure that preservation activities can be undertaken in a timely and organized way to best ensure data longevity;
iv. Limit the cost of storing and managing research data;
v. Ensure that legal obligations for research data storage and access are discharged.

Whyte and Wilson (2010) posit that appraisal and selection policy should be developed and implemented in conjunction with relevant stakeholders to ensure appraisal and selection process remains objective. Research data should be appraised, selected and disposed of in an organized, regular and documented fashion.

c) Data disposal

Data that is not selected for retention should be disposed of in accordance with the appraisal policy. Destruction may be appropriate and should be undertaken securely to ensure that personal or sensitive information cannot be accessed by unauthorized persons (British Standards Institute, 2009).

d) Data description

The description of research data is the 'backbone of research data curation' and it ensures that data can be discovered, identified, managed and retrieved (Treloar and Wilkinson, 2008). Data curation being imperative in RDM, requires proper and unequivocal data description through use of metadata standard that support some or all functions of research data such as, descriptive metadata, technical metadata, administrative metadata, user metadata and preservation metadata. Riley (2009) emphasizes that inadequate or incomplete metadata can make digital materials unusable in a short space of time as it may become undiscoverable, or the context of its creation may be lost.

e) Data preservation

Long-term preservation of authoritative nature of data is necessary for it to remain authentic, reliable and usable. Data Preservation means protecting data in a secure environment for long-term access and reuse, regular auditing to guarantee its integrity, using appropriate metadata to ensure its discoverability and monitoring to control access to meet privacy, licensing and intellectual property restriction (National Library of Australia, 2003; Treloar and Wilkinson, 2008).

f) Data access

The aim of data curation is the access, use and reuse of research data in addition to which consideration should be given from the conceptualization to access stage; who or what the data is for. Moreover, analysis of the research data should also be undertaken to ensure that relevant legislations regarding freedom of information and data protection are adhered to when allowing others to access it (Higgins, 2012). Therefore policies should be developed that establish who will be authorized to view, edit, download, upload or reuse the research data. Higgins (2012) further emphasizes that the provision of search and discovery tools should be appropriate and up to date for the designated community and considerations should be given to how searches will be undertaken and how data will be delivered.

g) Data use, reuse and sharing

Sharing data can also be used to advance the original research or another line of inquiry. Witt (2008) explains that preserving and sharing existing datasets could enable them to be reused instead of incurring the expense of generating new data from scratch. Funding agencies like NIH) and National Science Foundation (NSF) require deposits of publications derived from the research that they sponsor into open access repositories. In the same vein they also require a grant proposal including data management plans that address preservation and open access to the data generated (NIH, 2008; NSF, 2007b; Heidorn, 2011; Lyon, Patel, and Takeda, 2014).

h) Data transformation

Data transformation involves creating new data from the original material data, or turning the data into a different format. Higgins (2012) noted that transformed datasets can perform a variety of roles, such as verifying the results obtained from the analysis of the raw data and forming the basis for further experimentation. Therefore, the transformation of data finds the curation and research lifecycle restarted through the creation of new derived datasets and the imaginative reuse of data to underpin new research.

Data curation lifecycle can only exist within an institutional framework that has the personnel with the correct skills to perform each of the steps on a regular basis. The DCC lifecycle model therefore help curators understand the processes involved in developing curation and preservation methodologies for their research institutions (Palathingal, Dascalu, Harris and Varol, 2015).

2.2.2.2 Relevancy of DDC model to this study

Data curation is an effective way for addressing RDM. Higgins (2008) emphasizes that DCC lifecycle model advocates for maintenance of authenticity, reliability, integrity and usability of digital material which in return ensures quality of RDM. This view is particularly important for RDM in Kenya’s agricultural research institutes in view of the fact that the agricultural research data curated should maintain its relevancy over time by upholding its authenticity, reliability, integrity and usability.

Higgins (2008) further explains that DCC lifecycle model allows data curation activities to be planned at different levels of granularity which includes defining roles and responsibilities; building frameworks of standards and technologies; and ensuring that processes and policies are adequately documented. Moreover, Pennock (2007) opined that digital research data must not only be collected or created, but also properly managed, stored and preserved in order to maximize the initial investment and ensure that information remains reliable and available for users for as long as is deemed necessary. This is even more important given that agricultural institutes are investing in cyber infrastructure in order for research data curation to promote management, access, sharing, and reuse.

The University of Edinburgh (2016) asserts that data curation of agricultural research data remains important because it will allow data retrieval, avoiding unnecessary duplication, validating results, if required, and complying with funder's mandates. Consequently, Heidorn (2011) notes, those agricultural research institutions must curate data to protect and disseminate the intellectual capital of institution/society which is critical to the scientific and economic development of a country. Shakeri (2013) used the DDC model to investigate data curation perspectives and practices of researchers at Kent State University. Similarly, Heidorn (2011) used the model to investigate the emerging role of libraries in data curation and e-science.

2.2.2.3 Gaps in the use of the DDC model

The Model is outstanding in support of data curation (research question two), for agricultural research data and provides:

a) Data (digital objectives and databases)
b) Full lifecycle action (description and representation information; preservation planning; and community watch and participation)
c) Sequential actions (capture; appraisal; description; preservation; access, use and reuse; and transformation)
d) Occasional actions (dispose; re-appraisal; and migrate)

However, DDC operates in isolation of the institutional capabilities which are crucial in RDM for Kenya's agricultural research institute, for example, technical infrastructure, skills and training, collaborative partnerships and legal and policy issues. While the DCC lifecycle model provides a high-level view, it can be used in conjunction with relevant reference models, frameworks and standards to help plan RDM (Higgins, 2008; and Denzin & Lincoln, 2005).

2.2.3 Justification of using CCM Framework and DCC Lifecycle Model

This study adopts two theories namely CCM framework (Lyon et al., 2012) and DCC lifecycle model (Higgins, 2008) as a theoretical lens in addressing the research question. Additionally, there is an element of pragmatism ontology in adapting ideas from a range of theories as Creswell (2009) alludes that mixed methods research may both test and generate theories thus the study as adapted mixed method epistemology resulting in the use of two theories.

The study is underpinned by the CCM Framework and DCC Lifecycle Model. The CCM Framework addresses the community/institutional capabilities which incorporate skill and training, technical infrastructure, legal and policy issues, collaborative partnerships and openness which contribute a lot to RDM practices. DCC Lifecycle model equally prescribes the activities to be done in RDM such as data curation involving capture, appraisal, preserve, access and re-use. These activities are the core functionalities in RDM. The two theoretical models are used to address data curation and community/institution capabilities which are central theme to RDM. The CCM framework and DCC lifecycle model were chosen for its ability and elegance to explain a social phenomenon moreover the quality of a theory is judged by its explanatory power, its predictive power and its scope (Schoenfeld, 1998; Vithal, Jansen, & Jansen, 2013).

2.3 Other RDM models

There are other RDM models such as A Library-Oriented Model of Institutional RDM (Pinfield, Cox and Smith, 2014) and The Open Archival Information System (OAIS) Functional Model, (Laughton and Du Plessis, 2013).

2.3.1 A Library-Oriented Model of Institutional RDM

A Library-Oriented Model of Institutional RDM (Pinfield, Cox and Smith, 2014) was created with the library's perspective in mind; however it is used beyond the confines of the library. The model is intended to address 'The Who?', 'Why?', 'What?' and 'How?' of RDM, particularly in relation to the library's involvement. The model addresses the question, 'Who is involved in institutional RDM?' The RDM program components in the model address the question, 'What does an RDM program in an institution consist of?' The institutional drivers address the question, 'Why might a program be carried out?’ The influencing factors address the question, ‘How will the program be shaped?’ The following components form the different sectors of the models:

a) Stakeholders, (Who?): research support services, IT services, Library, other support services and senior university managers.
b) RDM program components, (what?): strategies, policies, guidelines, processes, technologies, services.
c) Drivers, (why?): storage, security, preservation, compliance, quality, sharing, jurisdiction
d) Influencing factors, (how?): acceptance, cultures, demand, roles, governance, politics resources, skills, and communications.

In contrast, the model faces limitations such as:

a) Scaling and operationalisation of activities,
b) Ensuring that all the different components of the program are all developing at a consistent rate,
c) Levels of compatibility and integration are a challenge.

Whyte (2014) is of the view that this model is not well developed in the treatment of the diverse disciplinary culture of the research community; therefore A Library-Oriented Model of Institutional RDM cannot underpin the study because its theoretical lens does not focus on the research questions of the study.

2.3.2 The Open Archival Information System (OAIS) Functional Model

The Open Archival Information System (OAIS) Functional Model (Laughton and Du Plessis, 2013) on the other hand consists of two sub-models; the first being the information model which deals with metadata that are used to preserve and access items in an archives. The second is the functional model that defines functions that are necessary for data curation which include ingesting, archival storage, data management, administration, preservation, planning and access functions (Laughton and Du Plessis, 2013).

Even though the OAIS Functional Model is a well-designed, model, it does have several shortcomings. According to Laughton and Du Plessis (2013), the OAIS functional model would be more effective if it had included a phase prior to the Ingest function and Pre Ingest. Pre Ingest could address the methods used for data collection with an aim of ensuring quality, understanding and accessibility of data (Laughton and Du Plessis, 2013). Nevertheless, the model does not have a theoretical lens that focus on any of the research questions of the study and as such, it is best suited for investigating research problems in archives and records management.

Literature will be reviewed on related areas of the variables mentioned so as to widen the scope of RDM.

Summary

This chapter presented theories underpinning the study and to understand the study, the researcher used four theories in an attempt to uncover the variables deemed necessary in the study. In reference to RDM in agricultural research institutes, CCM framework and DCC lifecycle model formed the theoretical foundation of the study. The mentioned theories offered a holistic approach to RDM which is seen as a continuous process in the lifecycle of RDM in agricultural institutes due to access, use, reuse, sharing and transformation. The mentioned theories underpinned the study in addressing data curation and community/institution capabilities which are central theme to RDM.

The chapter discussed relevant variable that were derived from the two theories that only underpinned the study. Further, relevancy and gaps of the theories informing the study were discussed including its justification in its use.

A Library-Oriented Model of Institutional RDM and the OAIS functional model were discussed in reference to RDM in research institutions. Though relevant to the study, they were not used to underpin the study because they were not well developed in the treatment of the diverse disciplinary culture of research community.

The following variables will form the thrust of the literature review in the next chapter:

- Legal, policy and regulations for RDM
- Data curation (capture, appraisal, describe, preserve, access, share and reuse)
- Knowledge, skill and training requirements needed for RDM
- ICT infrastructure for RDM
- Collaborative partnerships in RDM

Literature will be reviewed on related areas of the variables mentioned so as to widen the scope of RDM

CHAPTER THREE

LITERATURE REVIEW

3.1 Introduction

According to Mugenda and Mugenda (2003) literature review is the systematic identification, location and analysis of documents that contain information that may be related to the problem that is being studied. Similarly, literature review is an interpretation of a selection of published and /or unpublished documents available from various sources related to the area of study that optimally involves summarization, analysis, evaluation and synthesis of the documents (Onwuegbuzie, Collins, Leech, Dellinger and Jialo, 2010). It can be noted that literature review plays a crucial role in preparing and orientating the researcher with regard to the raging debates taking place in the field in which he or she is about to conduct research.

According to Bloomberg and Volpe (2008: 26) literature review sets the broad context of the study, clearly demarcating what is or is not within the scope of the investigation, and justifies those decisions. The aim is to gain insight into and knowledge about what has been done, by whom, where, why, how and with what kind of results. De Los Reyes and Kazdin (2008) state that literature review places the current research in its historical context; it describes the background to the study and the relationship between the present study and the previous studies conducted in the same area thus identifying trends and debates in the existing literature. More importantly, literature allows the researcher to gain insight into what has been achieved in the past and also offers the researcher new perspectives that have not been adequately covered by previous researchers.

Literature review is important (Cooper, 2011:20) because it can:

1) Integrate what others have done and said;
2) Criticize previous scholarly works;
3) Build bridges between related topics; and
4) Identify the central issues in a field.

To this end, Pautasso (2013) buttress that the sole purpose of conducting a literature review is to seek ideas that support and defend your own views and also advance your arguments, which to a large extend are posited as answers to the research question or problem. The rationale of literature review in this research is therefore to justify why the study is important. This literature review would support and defend the researchers' own arguments, claims and efforts. (Mathipa, 2015).

The purpose of this study is to examine research data management in Kenya's agricultural research institutes with a view to proposing interventions to improve management, sharing and reuse of agricultural research output.

Empirical and conceptual literature reviewed in this chapter is obtained from books, journals, theses, conference proceedings, databases, among others. Kothari (2004) points out that there were two types of literature: the conceptual literature concerning the concepts and theories and the empirical literature which discuss studies related to the variables of the current study.

The chapter is organized around themes of research questions, key variables of the underlying theory and broader issues on the research problem. Thematic areas from the research questions include: Research Data Management practices (RDM); legal, Policy and regulations affecting RDM; research data capture, appraisal, description, preservation, access, and reuse; Knowledge, skills and training requirements of RDM; ICT infrastructure for RDM; and Collaborative partnerships influencing RDM. In addition, key variables from the underlying theory are RDM advocacy and data literacy, RDM security and a data management plan. Moreover, the broader issues around the research problem in this study include scholarly communication, agricultural research data, and data repositories, library and IT departments. Within each theme, the international context is reviewed followed by regional, national and local contexts.

3.2 Research Data Management (RDM)

In the wake of a 'knowledge society', and the 'knowledge economy', knowledge and information are considered public goods and essential ingredients for development. In this regard, RDM is now a widely accepted practice that adds value to organizations in the form of new information and knowledge and therefore RDM is currently the central focus in institutional planning, management and innovation (Hoq and Akter, 2012; Sörlin and Vessuri, 2007).

The availability of huge amounts of new data, often referred to as a data deluge, has the potential to revolutionize the method in which research is carried out, and this can consequentially result in new e-infrastructure (research data infrastructure) developed for purposes of harnessing the accumulating data and knowledge produced by research communities through data curation, open access, shared and reused research data. RDM brings benefits to researchers and research institutes in many ways (Lewis, 2010; Piennar, 2010) such as:

a) Meeting research funders requirements;
b) Ensuring research integrity and replication;
c) Ensuring research data and records are accurate, complete, authentic and reliable;
d) Increasing research efficiency;
e) Saving time and resources in the long run;
f) Enhancing data security and minimize the risk of data loss;
g) Preventing duplication of effort by enabling others to use your data; and
h) Complying with practices conducted in the research institutes in the county.

The study by Van den Eynden (2013); and Cox and Pinfield (2014) on the UK Data Archive revealed that the over forty (40) years' experience in selecting, ingesting, curating and providing access to social science, has enabled sharing of the data with organizations in the same discipline area and has thus played a big role in RDM in UK (UK Data Archives, 2017). According to the University of Sheffield (2016), good RDM ensures that data produced or used during research activities is preserved, made accessible for use and reuse, managed over time and disposed of according to legal, ethical and funder's requirements. The debates about RDM have gained great impetus because of the large amount of research data that is generated because of the influence of ICT. Consequently, RDM has emerged as a strategic priority for research institutes (Pryor, 2012; Whyte and Tedds, 2011).

Organisation for Economic Co-operation and Development (OECD) (2007) points out that the principles and guidelines of open access is meant to apply to research data, whether already in existence or yet to be produced and supported by public funds for the purpose of accessibility. These principles and guidelines include openness, flexibility, transparency and legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability; and sustainability.

Africa has made strides in RDM. Van Deventer and Piennar (2015) state, that South Africa is leading the cluster of African countries in embracing RDM. Lötter (2014) and Fernihough (2011) similarly assert that Data Intensive Research Initiative of South Africa (DIRISA) is one of the initiatives aimed at promoting RDM in that country. The main purpose of DIRISA is to: coordinate data science capacity development; fund capacity development; promote and support priority research but with caveats of data stewardship planning and capacity building; provide services and research environment; promote data contribution and adoption of open standards and open data; and support data stewardship in a federated context (Vahed, 2014). DIRISA has made South Africa leap a head in terms of RDM in all different sectors of the economy, especially in academic, health and research institutions.

In Kenya, there is some attempt albeit limited to promote RDM, for instance, Kenya's public health officials and researchers are often faced with the complex reality of making health-related decisions in the absence of coordinated research data and functional data management systems responding to health related questions is a challenge (Family Health International-Kenya, 2005). Kiplang'at (2004) asserts that agricultural research in Kenya contributes to increased agricultural productivity in terms of yields, quality, profitability and sustainability of enterprises.

3.3 Agricultural research data

Investment in agricultural research has the potential to promote economic growth, poverty alleviation and improvement in food security. This can only be attained when agricultural research data is curated, accessed, shared and reused. Agricultural research data refers to data in particular facts or numbers generated from experiment, measurements, survey results, observations, interview recordings and images from agricultural activities (European Commission, 2016). Users of agricultural research data can access, mine, exploit, reproduce, use and reuse under the terms and conditions set out in the legal or institutional policies.

The processes by which agricultural research data are generated and shared are being transformed and reinvented by the application of ICTs. Moreover, Edge, Rudgard, Maru and Porcari (2011) assert that these changes (use of ICTs) are providing opportunities for agricultural researchers worldwide to interact and jointly develop, share and reuse research data to enhance agricultural production by disseminating information and knowledge generated from research through extension workers to farmers.

Agricultural data repositories present opportunities to capture and share a range of different types of agricultural research content. Similarly, the recent rapid spread of Web 2.0 tools, intranets, extranets, portals and related communication activities, has opened up low cost routes to capture and disseminate research data in more informal ways (Edge et al., 2011). It is therefore within this context that it has become increasingly clear that agricultural research data that remain to be captured, disseminated and communicated are generally invisible and for that reason wasted investment.

A study undertaken by Lord et al., (2004); Awuor, Rabah and Kimeli (2013); and Mcnamara (2009) concerning the role of ICTs on agriculture, has established that ICTs has made huge impact on agricultural practices by making it increasingly data-intensive as a result of the proliferation of digital technologies, instrumentation, and pervasive networks through which data are collected, generated, appraised, preserved, shared and reused. The enormous growth in the availability and utility of agricultural research data is increasing scholarly research productivity, accelerating the transformation of research outcomes into products and services, and enhancing agricultural productivity and food security (NSF, 2007a).

The Food and Agriculture Organization (United Nations) (FAO), the International Fund for Agricultural Development (IFAD) and the World Food Program (WFP) (2012), have established that agricultural research and extension are becoming increasingly globalized, focusing on technologies that are knowledge-intensive, while requiring RDM to facilitate access, sharing and reuse to enhance agricultural productivity and food security. On the contrary, Pauw and Thurlow (2011) point out that, small-scale farmers in developing countries face challenges in accessing and properly utilizing knowledge and information generated through agricultural research data due to high illiteracy levels and poor infrastructure. FAO, IFAD and WFP (2012) note that governments should enhance support to small-scale farmers in developing countries by ensuring high quality agricultural research data are accessed, shared and reused by target groups. Supporting small-scale farmers in utilizing agricultural research data has a number of benefits, including meeting food demand, as well as enhancing food security and nutrition in rural and urban areas.

Studies by United Nation Development Program (UNDP), (2012); Edge et al., (2011); and Stienen, Bruinsma and Neuman (2007) on agricultural research data and its impact on smallholder farmers in developing countries established that low investment and lack of coherence in technical infrastructure, skills and training, legal frameworks and policies, and collaborative partnerships in agricultural research data have had minimal beneficial support from digital open environment thus undermining the efforts to combine and share agricultural research data, information and knowledge between countries and continents. Agricultural research data is increasingly recognized as a vital resource whose value needs to be preserved for future research. This places a huge responsibility on agricultural research institutes to ensure that their research data is managed in such a manner that the ultimate intended beneficiary accesses research data and are protected from substantial reputational, financial and legal risks in the future.

3.4 Research Data Management (RDM) in agricultural research institutes

RDM is crucial in facilitating data sharing and ensuring the sustainability and accessibility of data in the short and long term. Van den Eynden (2013) acknowledges that data created from agricultural research are valuable resources that can be used and reused for future scientific and educational purposes; in addition, sharing data facilitates new scientific inquiry, avoids duplication in data collection and provides rich real-life resources for education and training. In this respect, agricultural research data can be used and reused for future research if shared, managed well, properly preserved and made available.

Agricultural researchers are challenged by the strategies of collecting and managing data in the short and long term, as well as by their reuse. Without good management and organizational practices, research data are difficult or impossible to access. RDM often requires vast storage space and good preservation strategies (Lord and Macdonald, 2003). Similarly, digital data require appropriate management strategies including a well-developed curation plan to promote access, sharing and reuse. Consequently, Bracke (2011); and FAO, IFAD and WFP (2012) point out that agricultural research data is required to be managed on appropriate electronic formats which should be supported by investment in capacity, capability and e-infrastructure.

RECODE (Recommendations for Open Access to Research Data in Europe) Project Consortium (2014) acknowledges that RDM costs in agricultural research institutes include those for data capturing, appraisal, preservation, accessing, sharing, training, technical infrastructure, policy formulation and out-reach programs. Therefore, developing sustainable funding models and establishing collaborations to facilitate RDM should be addressed.

Why manage Research data? Overwhelmingly, research data outputs are produced at an increasingly rapid pace while the amount and variety is vast. Pryor (2012) and Van Wyk and Van der Walt (2014) enumerate reasons why research data needs to be managed namely to: ensure research integrity and replication; prevent duplication of effort by enabling others to use existing data; ensure research data and records are accurate, complete, authentic and reliable; as well as to enhance data security and minimize the risk of data loss

Physical, network and computer systems as well as file security are needed to ensure security of data and prevent unauthorized access, changes to data, disclosure or destruction of data (Van den Eynden, Corti, Woollard, Bishop and Horton, 2011). Data security arrangements in agricultural research institutes need to be proportionate to the nature of the data and the risks involved at all stages, including data destruction stage.

Good RDM in agricultural research institutes contribute to well organized, documented, preserved and accessible research data while their accuracy and validity is controlled at all times. (Van den Eynden et al., 2011). A study done by Van Wyk and Van der Walt (2014) conducted interviews on 52 respondents with regard to RDM at University of Pretoria and the findings revealed that successful RDM should be supported by appropriate but clear RDM policies as well as establishment of a central RDM office, data repository and installation of necessary IT infrastructure for RDM and this should be replicated in agricultural research institutes.

Van Deventer and Pienaar (2015); Olum (2013) and UNEP (2015) in examining RDM in Africa, established that research data are scattered across computers or other filing systems, differ in formats or in software versions, or are lost as researchers transition to other projects. However, Alila and Atieno (2006) suggest that agricultural research institutes should consider a data curation plan, while educating researchers on better data organization and management practices.

A presentation by Mutula (2014) on the impact and development of RDM at the University of KwaZulu-Natal sums up the critical issues affecting RDM to include:

1) Sheer amount of data generated that has to be managed;
2) Complexity of the research process;
3) Heterogeneity of data and formats;
4) Varied technologies for storing and managing data; and
5) Ethical aspects of research.

Additionally, Mutula (2014) recommends actions needed to improve RDM:

1) RDM road map (RDM planning services, active data infrastructure, data management advocacy, data management capacity building, data recovery and backup services);

2) Development of appropriate RDM policy; and
3) Development of RDM capabilities

Alila and Atieno (2006) assert that agricultural policy in Kenya does not clearly define RDM in agricultural research institutes to give effect to guidelines on human capacity and technical infrastructure required for access, sharing and reuse of research data.

3.5 Legal, policy and regulations affecting RDM

Vast quantities of agricultural research data are being produced in a wide variety of forms and at a rapid rate in Kenya's agricultural research institutes, hence creating volume and variety challenges in RDM. Pinfield, Cox and Smith (2014); Whyte and Tedds, (2011); Antell, Foote, Turner and Shults (2014) assert that in order to solve challenges of research data management, a need exists to develop legal and policy framework, infrastructure and services in research institutions to manage data with the aim of assisting the creation, collection, manipulation, analysis, preservation, accessing, sharing and reuse of research data.

In essence, RDM legal and policy framework should respond to a number of RDM drives such as: storage, security, preservation, quality, compliance, sharing, and jurisdiction in order to enhance management, sharing and reuse of agricultural research output (Pinfield, Cox and Smith, 2014; Higman and Pinfield, 2015). Legal, policies and regulation framework affecting RDM is informed by DCC lifecycle model which defines process and policies at every stage of the lifecycle whereas CCM framework maps the entire legal, policy and regulatory framework of data curation and RDM capabilities.

3.5.1 RDM legal framework

RDM legal framework is an opportunity to start developing best practices in RDM in a proactive manner reflecting on the mandate of agricultural research institutes. Smith (2014) in a study on data governance argues that legal environment surrounding research data lags behind hindering the ability to develop best practices for data management, sharing and use. Moreover, RDM legal environment includes laws, regulations, and policies associated with data, as well as strategies for data quality control and management in the context of agricultural research institutes. In this respect, RDM legal framework ensures that data can be trusted through facilitation of RDM governance by adoption of appropriate technical standards, practices and architecture that will necessitate management, sharing and reuse of research data.

A study by Fitzgerald, Pappalardo and Austin (2008) on RDM legal and policy guide advances the view that RDM legal framework should be comprehensive, coherent and precise by covering issues such as copyright protection, moral rights, patents, privacy, confidentiality, internal guidelines and policies, and contract. Similar sentiments are shared by Patel (2016) in a study that adopted qualitative and quantitative epistemologies in examining RDM conceptual framework where he recommended that before embarking on a journey with RDM, there has to be an institutional RDM legal and policy framework that clearly spell out the following in the context of RDM: purpose, scope, applicability and guidelines to the data contributors, licensing, metadata, data classification, copyright agreements and conditions, terms and conditions of the use of data, protection of confidentiality of sensitive data, protection of data against security breaches and intellectual property concerns. In addition, Karick (2014) points out that RDM framework should clearly state ownership and rights associated with research data in order to minimize disputes and provide accountability for research data.

3.5.2 Policy and regulations affecting RDM

Policy and regulations affecting RDM must be developed to address data curation, quality and security as well as ethical requirements, human resource capacity, technical infrastructure and collaborative partnerships at every stage of RDM. This should result in the development and operationalisation of an effective and efficient collection, appraisal, preservation, access, reuse and sharing of research data (Cox and Pinfield, 2014). Mullins (2014); Higman and Pinfield (2015); Erway (2013); Keralis, Stark, Halbert and Moen, (2013) adopted quantitative and qualitative epistemologies in examining a range of RDM policies covering funders of research projects, preservation, open access, data curation, sharing, reuse, and collaborations. These studies found that many research institutions were developing new institutional RDM policies with the involvement of stakeholders in the formulation, establishment, documentation, implementation, maintenance and promulgation of policies, procedures and practices for RDM. In addition, qualitative analysis conducted by Pinfield, Cox and Smith., (2014) showed that the policy development process typically involves consultation with RDM stakeholders across the institution. Such process is often iterative in nature, and policymakers commonly adapt elements from other institutions to suit their own needs.

According to RECODE Project Consortium (2014), the allocation of responsibilities to each player in the RDM should be spelled out in the policy which should also be flexible enough to accommodate changes that suit the needs of researchers and keep pace with technological developments. A study done by Bracke (2011) on curation of agricultural data established that there are often no data standards within a discipline and researchers often have very idiosyncratic research practices which increases complexity in agricultural research data. Mossink, Bijsterbosch, and Nortier, (2013) and MacKenzie (2014) on the other hand found that whereas most funders of research projects will have policies covering data management, there is inadequacy of data management policies covering training and support, security, data curation, preservation, reuse and open access. An RDM study conducted through an online survey of 145 research funders, national bodies and research institutions by Mossink, Bijsterbosch, and Nortier, (2013) focusing on Support Infrastructure Models for RDM (SIM4RDM), underscored the fact that a number of research institutions with RDM policy is growing, however there is a need for institutional policies on RDM to have the following elements:

a) Responsibilities and roles;
b) Access and reuse of data;
c) Long term preservation/curation;
d) Security;
e) Open accessibility and availability of data;
f) Protection of legitimate subjects of research data (embracing informed consent, anonymity and confidentially).
g) Provision of mechanisms for storage, back up, registration;
h) Training, support, advice;
i) Copyright and intellectual property issues (copyright, patents, trademarks and design rights);
j) Embargo period for open accessibility;
k) Methods used to share guidelines/restrictions on data;
l) Destruction of records;
m) Removal or transport of data;
n) Preferred Licenses for data.

The study recommended further that research institutions should build a trustworthy research data infrastructure and create workflows for data publishing and archiving by raising awareness amongst researchers regarding important research data related aspects. Similarly, the study concluded that there is a need for policies and guidelines to govern ownership of data created within the agricultural research institutes as this is a problematic area (Mossink, Bijsterbosch, and

Nortier, 2013). In addition to the elements listed by Mossink, Bijsterbosch, and Nortier, (2013) on what the policy should entail, Smith (2014) posits that RDM policy and regulation framework should address issues such as copyright, data licensing, data security, data privacy, and mind-set (researchers to accept the idea of their data being made available for reuse). Therefore, policy development is a cross-institutional process and by initiating the conversation about RDM policy, it should be an all-inclusive with the goal to facilitate effective and efficient management, sharing and reuse of agricultural output (Erway, 2013).

Nevertheless, Africa in a small way, has embraced RDM policies and related procedures in research institutions. Kahn et al., (2014) and Woolfrey (2014) singled out University of Pretoria as having formal RDM policy, while UCT, Stellenbosch University, UNISA, University of KwaZulu-Natal (UKZN) and University of the Witwatersrand lacked RDM polices but have related policies such as codes of research practices and research ethics. Furthermore, Lötter (2014) did a study on RDM position in South Africa and reported that more universities are now getting involved in RDM to varying degree, for instance, analyzing researcher's awareness of RDM, engaging with management about RDM and taking part in learning activities related to RDM through conferences and workshops.

In the same vein, Mukiibi (2016) examined policy guidelines and research working documents of research institutions in Uganda, and reported that they lacked the aspect of RDM or were silent on and about RDM. This is an indication that no work has been done on the subject of RDM in Uganda despite great strides that have been achieved in some African nations. Mukiibi (2016) also pointed out that Health and Medical related research have research policy guidelines that provide for preservation of research data for only five years after the research has been completed and permanently disposed of it thereafter. Unfortunately, Uganda has not yet moved with the trend and therefore remains behind in enforcing RDM.

While most research institutions tend to have overarching RDM policies they lacked specific policies governing every stage of RDM cycle. Most importantly, RDM policies in most agricultural research institutes need to be reviewed to mitigate existing sociological and cultural barriers to data sharing and access and to bring them into accord across programs and ensure coherence (NSF, 2007a).

In sum, studies on policy and regulation development appears to be most effective when it is undertaken iteratively and in close cooperation with researchers across a range of disciplines as well as other stakeholders from agricultural research institutes.

3.5.3 Quality assurance and control measures of research data

Quality assurance and control measures of RDM remains an integral part of the process and takes place at various stages of the lifecycle of research data but often neglected as a role in ensuring high quality research data. In this respect, proper protocols and methods must be employed to ensure that research data are properly collected, handled, processed, used, and maintained at all stages of the scientific data lifecycle in order to guarantee high quality research data. Quality control measures in RDM is defined as methods and procedures implemented to ensure that research data are collected, managed, and utilized with accuracy and precision while quality assurance in RDM can be defined as a set of activities designed to ensure that a product or service meets specified requirements (project audits and process checklists) (UK Data Archive, 2017; Reisch and Webster, 2004; United States Geological Survey, 2017). The main difference between quality assurance and quality control measures of RDM is the fact that quality assurance is process oriented, focusing on building quality to prevent defects and make sure that researchers and other RDM stakeholders are doing the right things, the right way. Whereas quality control measures of RDM is product oriented, focusing on testing for quality and making sure the results of what researchers and other data professional staff have done are what they expect (United States Geological Survey, 2017). All said, quality assurance and quality control measures of RDM is an integral part of all research and takes place at various stages, during data collection, data entry, and data checking (Van den Eynden et al., 2011).

Pandav, Mehta, Belle, Martin, Chandra et al. (2002) explain that quality assurance and quality control measures of RDM are based on the principle that these processes begin at data collection protocol and ends when data archiving is complete by building into the RDM system and carried out concurrently with other RDM activities. Further, Food Security and Nutrition Network (n.d) proceed to illustrate that research data possess seven (7) key high quality assurance and control attributes namely: validity; reliability; precision; completeness, consistency, integrity; and timelines. In order to achieve the attributes above, the following should be taken into consideration for successful quality assurance and quality control measures in RDM:

i. Adequate staff capacity, supervision and accountability,
ii. Complete documentation of processes/ protocols readily available to collectors and processors,
iii. Routine cross checking-mechanisms,
iv. Clear strategy to respond to problems, and
v. Adequate financial and logistical resources to ensure timely performance.

UK Data Archive (2017) in a study on quality assurance explained that during research, the quality of research data collection methods used, strongly influence data quality, and documenting in detail how data are collected provides evidence of such quality. In addition, when agricultural research data are digitized, transcribed, entered in a database or coded, quality is ensured by standardized and consistent procedures for data entry with clear instructions. Consequently, research data need to be checked continuously on a routine basis for editing, cleaning, verifying, cross-checking and validation.

Quality assurance and Quality control of RDM should be paramount in Kenya's agricultural research institutes because it ensures the quality of agricultural research data before it is collected, entered, or analyzed. To ensure quality assurance and quality control in RDM, Kenya agricultural research institutes are currently ISO (International organization for Standardization) compliant implying that research data capture, appraisal, description, preservation, access, reuse and sharing are strongly influenced by ISO procedures and standards. The adoption of ISO 9001: 2008 in Kenya agricultural research institutes has boosted the quality agricultural research data because the institutes have been compelled to come up with Standard Operating Procedures (SOPs) which are in-line with KALRO, ISO 9001:2008 to assist in data curation. In order to have a robust quality assurance and quality control for RDM, policies, regulation and strategies need to be put in place to enforce quality of research data.

To bridge the gap in literature, the first research question (see section 1.5.1.1) sought to establish the availability or absence of legal, policy and regulations affecting RDM.

3.6 Data curation

Data curation as one of the key aspects of RDM involving research data capture, appraisal, description, preservation, access, reuse and transformation. Data curation is defined as managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose and available for discovery and use (Lord et al., 2004; Walton, 2010; Laughton and Du Plessis, 2013; and Bracke, 2011). Data curation is the active and on-going management of data throughout its lifecycle of interest and usefulness to research to enable data discovery and retrieval, maintain quality, add value and provide for reuse over time (Palmer, Cragin, Heidorn, and Smith, (2007). Data curation is underpinned by DCC lifecycle model which is committed to promoting a lifecycle approach to research data to enable their successful curation, reuse and sharing.

Heidorn (2011); Antell et al., (2014); and Karasti, Baker and Hakola, (2006) have discussed extensively the subject of data curation using the Data Curation Centre (DCC) Lifecycle Model. These studies that have largely assumed qualitative epistemologies and descriptive approaches emphasize the need for more efforts to be directed towards understanding practices of data curation and stewardship. Heidorn (2011) advances the view that there is a large volume of data currently not being curated adequately. However agricultural research institutes and funding agencies are encouraged to recognize the importance of curating data to protect and disseminate the intellectual capital of society.

Heidorn (2011) conducted a study on the role of ICTs on data curation and established that data curation, has enhanced use of computers and telecommunications which have revolutionized the methods for collecting, storing, analyzing, nurturing, disseminating and using of research data. The proliferation of readily accessible data capture devices available to researchers has led to a vast increase in the amount of data being produced through research. While such devices may have made data creation easier, the volume and distributed nature of this data has led to increased challenges surrounding collecting, curating, preserving and reuse in terms of RDM illiteracy, data storage and mode of sharing research data (Jahnke and Asher, 2012).

Data curation is of great interest to governments, funders, agricultural research institutes and researchers. Whyte and Allard (2014) acknowledge the challenges in data curation with regard to lack of legal framework, standards or procedures to reference and define mandatory guidelines when curating data. Research trends in data curation as reported by Weber, Palmer and Chao (2012); Carlson and Leiter (2009); and Kim, Addom and Stanton (2011) highlight important data curation aspects which include:

a) Developing interoperable standards for describing and interchanging datasets;

b) Need for data curators to participate in the formulation of data privacy and ownership policy;
c) Need for a workforce skilled in data curation practices; and
d) Development of professional education standards guided by data-driven research agenda.

In reviewing literature on data curation, Palathingal et al., (2015) concluded that a global trend on curating agricultural research data in the emergence of data-intensive research call for a well-designed technical infrastructure, trained human capital, policies and procedures at every stage of data curation and collaboration among agricultural research institutes.

3.6.1 Data capture

Data capturing entails designing research, data management plan, plan consent for sharing, locate existing data, collect data (experiment, simulate, observe), capture and create metadata (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). Australian National Data Service (n.d (a)) defines data capture as the process of collecting data which will be processed and used later to fulfill certain purposes. The capture or create stage of the curation lifecycle focuses on ensuring that data captured is fit for purpose and ready for curation.

Agricultural research data with good metadata attached at the point of capture can expedite data sharing, publishing and citation. Metadata capture is of great value simply because the more information there is about data, the greater the value of the data whether automatic or manual. Similarly, Jones, Guy and Pickton (2013) point out that at the capturing stage file naming, versioning and structuring of files needs to be performed to ensure ease of accessing data when needed, bearing in mind both the short-term and long-term. Consequently, Australian National Data Service (n.d (a)) buttress that data capture tools should:

a) Provide processes of organizing and structuring data files;
b) Have data validation components to ensure that captured data meet required types and ranges;
c) Enable open and flexible formats where good conversion tools exist; and
d) Allow data to be moved to its destination efficiently and with high quality.

Higgins (2012) advances the view that relationships developed between the researcher and the information professional is very crucial in planning research data capture, since the former is concerned with capturing data, while the latter is concern with receiving and ensuring data capture.

In South Africa, researchers are increasingly aware of the importance and value of capturing and sharing research data produced through public funding with regard to RDM policies (Kahn et al., 2014). Metadata should therefore be captured as closely as possible to the creation of the data and the process of capturing data should be compliant with privacy and ethical regulations.

3.6.2 Data appraisal

Data appraisal and selection are fundamental stages in data curation process entailing: selecting appropriate data as well as entering, digitizing, transcribing; checking, validating and cleaning data; anonymising data where necessary; describing data; and managing and storing data (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). Appraisal and selection is the process of evaluating research data in order to decide which to retain over the long term, which to retain for the meantime and which to discard (Higgins, 2012; International Standards Organization (ISO), 2001).

Due to the large quantities of agricultural research data, appraisal has been considered one of the most important functions in data curation as well as the RDM. However, there is need to adhere to documented guidance, policies and legal requirements in evaluating research data and selecting those requiring long-term preservation. Appraisal and selection policy need to ensure consistent, transparent and accountable decision making. Whyte and Wilson (2010) state that appraisal and selection policy must fit legal requirements, for example, relating to privacy and intellectual property rights, Public Records Acts, national data policies and codes of conducts adopted by the host institution or agricultural research institute or funders.

According to Niu (2014), the archives and records management community has created and accumulated rich theories and methods for appraising both traditional and electronic records which can be applied to other types of digital resources, such as digital publications, web resources and agricultural research data. Harvey (2006) on the other hand tried to create a framework for appraising/selecting research data by adopting some archival appraisal theories and methods and reviewing existing appraisal/selection policies for preservation and digitization. In this regard, Harvey (2006) created ten appraisal criteria including: value, physical condition, resources available, use, social significance, legal rights, format issues, technical issues, policies and documentation. Niu (2014) critiques appraisal criteria in Harvey's framework by revealing that it was not well thought out as it was mutually exclusive, for instance, value and significance are used as two separate appraisal criteria, although social significance is a kind of value.

To this end, Niu (2014) and Eaker (2016) came up with a framework that synthesized traditional archival appraisal methodologies and elements identified from existing appraisal policies. Moreover, Niu's appraisal/selection framework is intended to serve as high-level guidance for individual institutions to create their local appraisal/selection policies for research data. Niu (2014) and Eaker (2016) appraisal/selection criteria include:

a) Mission alignment: whether the resource supports the mission and falls within the scope of the collection policy of a preserver or the institution;
b) Value of digital resources: can either be primary or secondary value as Tibbo (2003); Schellenberg (1956); and UK National Archives (2012) point out; that primary values include administrative, fiscal, and legal value while secondary values includes evidential and information values. In correlation, values of resources/research data are affected by the following factors, authenticity, reliability, integrity, accuracy, usefulness, usability, accessibility, uniqueness, diversity and representativeness (National Archives and Records Administration (NARA), 2007; Inter-University Consortium for Political and Social Research (ICPSR), n.d; and Whyte and Wilson, 2010);
c) Cost: It could be costs in acquiring, housing, preserving, and processing the collection to make it accessible also assess whether value of the data exceeds costs;
d) Feasibility: Feasibility of preservation is often determined by the technical capacity which include metadata and documentation, file formats and protection mechanisms such as password, digital signatures or encryption (NARA, 2007).

Therefore, there is need for researchers, librarians, archivists and IT specialists in agricultural research institutes to conduct appraisal primarily to decide whether to acquire research data for permanent preservation or for a limited time period. Consequently, there is an added need to decide the retention periods as National Digital Stewardship Alliance (NDSA), (2003) notes:

> Appraisal as a process by which archivists and records managers assign administrative, legal, research, and historical value to records in order to determine retention period.

In light of this, the 'appraise and select' action is very imperative in Data Curation Centre (DCC) lifecycle model as well as RDM.

Whyte and Wilson (2010) observed that in agricultural research institutes, research data librarian and archivist should assume the role and responsibility of setting appraisal/selection criteria and appraisal policy in consultation with stakeholders, especially researchers and local data managers in order to know how data would be assessed and how they would increase their enduring impact.

3.6.3 Data disposal

Data that is not selected for retention should be disposed of in accordance with the appraisal policy. Disposal may include permanent archiving, redeployment, transfer of custody or ownership, or destroying research data (Higgins, 2012). When the specified period of retention has finished, researchers have a responsibility to dispose of research data in a secure and safe manner in accordance with documented guidance, policies and legal requirements. According to DCC lifecycle model (Higgins, 2008), research data which has not been selected for long-term curation and preservation should be disposed of in accordance with documented policies, guidance or legal requirements. University of Tasmania (2016) and Van den Eynden et al., (2011) suggest ways of disposing research data: one; shredding papers or use of secure destruction service bin; two, deleting or overwriting digital data; third, purging magnetic media through degaussing or destroying the physical media (CD-ROM, DVD).

Higgins (2012) also notes that destruction may be appropriate but should be undertaken securely to ensure that personal or sensitive information cannot be accessed by unauthorized persons. In this regard, agricultural research institutes should take data disposal with the seriousness it deserves by following the disposal policy.

3.6.4 Data description

Data description entails interpretation of data; derive data; produce research outputs; author publications; data anonymisation; data visualization; data validation; metadata creation; data verification; data interpretation and analysis; and prepare data for preservation (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). Research data description is the documentation that accompanies the research data which makes it discoverable and usable over time thus metadata standards exist to provide standardized descriptions, for example, Dublin Core, and computer software (The University of Western Australia, 2016). Furthermore, metadata assists agricultural research data and scholarly publishing in a way of discoverability of data, data identification, data association with publication and related datasets, quality assurance and validation of data. A similar sentiment is shared with Patel (2016) who points out that to facilitate accurate retrieval of data, a description and identification of the data is important and this can only be achieved through specialized metadata schema for data representation.

Proper description and documentation of research data allows users to understand and track important details, in addition to describing research data using metadata facilitates, searching and retrieval in data repositories. Regarding the contents of metadata in describing research data, Cornell University Library (n.d) gives an example of the content of metadata such as contact information, geographic locations, units of measure, abbreviations or codes used in the dataset, instrument and protocol information, survey tool details and much more. Such detailed metadata content facilitates data curation which results in quality RDM. Data description is imperative in agricultural research institutes in terms of fully describing research data for easy accessibility, searching as well as retrieval and preservation.

3.6.5 Data preservation

Data preservation entails migrating data to the best format and suitable medium, back-up and store data, create metadata and documentation, and archive data (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). Data preservation means securing permanent access to the original research data from the finished research project and general characteristics of data preservation is data accessibility to others for verification or for sharing or collaboration within the scientific community (Kruse and Thestrup, 2014). Subsequently, it is imperative that long-term preservation and protection of sensitive data are vital characteristics of data preservation actions in agricultural research institutes.

Furthermore, policies and requirements regarding where to deposit research data and the retention period should be clearly stipulated as well as address the gap between short-term access and long-term preservation with reference to the type of research data (RECODE Project Consortium, 2014). Activities that support the preservation process should be planned to include administrative procedures required before undertaking preservation activities and the technical requirements of preservation.

A survey carried out by Mossink, Bijsterbosch and Nortier (2013) on Support Infrastructure Models for RDM (SIM4RDM) in Europe revealed that institutional repositories were also deployed to store finished datasets, for example, the Netherlands, the UK and Finland have well-established data archived for storage. Furthermore, the selection of file formats for data storage plays a crucial role in determining the long-term accessibility and share-ability of agricultural research data. Western libraries (2012) is of the view that digital preservation ensures the on-going management of digital research data over time with a view to retaining their intellectual content, authenticity, and accessibility for a variety of uses.

Agricultural research data preservation is an ongoing process that should be planned throughout its lifecycle to include the following activities: validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures (Data Curation Centre, 2004b). These preservation actions should ensure that agricultural research data remain authentic and reliable while maintaining their integrity.

3.6.5.1 Data repository

Data repositories are another strategy that could facilitate preservation of agricultural research data to ensure that researchers, scholars and other stakeholders can be assured of availability, accessibility, preservation and dissemination of agricultural content (Decman and Vintar, 2013; and Yiotis, 2008). Agricultural data repositories are crucial in RDM because as an online archive they perform various roles including collecting, preserving, and disseminating digital copies of intellectual output.

An effective data repository is usually achieved through the collaborative works of librarians, information technologists, archivists, policymakers and agricultural research institute administration. Agricultural data repository normally performs two functions: first, to support agricultural research institute in relation to annual reports, research assessment exercise among others and second, showcase research output of agricultural research institute (Lynch, 2003). On their part, research funders and publishers require agricultural research institutes to deposit research data in certified and credited repositories, in an effort to secure the reusability and long-term preservation of research data. RECODE Project Consortium (2014) agrees that obtaining accreditation or certification to appropriate standards is a way for ensuring both the quality of data repositories and of the quality assurance process.

Data repositories have developed intellectual capability ratings of research institutions especially agricultural research institutes, these intellectual capability rating have associated benefits (Queensborough Community College, 2016) such as:

a) Opening up outputs of the institution to a worldwide audience;
b) Maximizing the visibility and impact of these outputs as a result;
c) Collecting and curating digital output;
d) Enabling and encouraging interdisciplinary approaches to research;
e) Providing access, use and reuse, and sharing information, knowledge and research data; and
f) Providing a workspace for collaborative research.

Good agricultural data repositories are goldmines for agricultural research institutes because they bring in the benefits of open access as well as enhance sharing and reuse of research data. On the contrary, Parker (2012) points out that there are complexities around clarity of ownership, description and preservation formats of research data in data repositories which results in some discontentment among researchers. In this regard, Amorim et al. (2015); Bush (2009); and Fary and Owen (2013) state that if a clear and articulated RDM policy is in place, then issues of ownership, storage, formatting, description, networks, and software will clearly be spelled out to allow fluent RDM.

Africa has embraced research data repositories as evidenced by a study done by Kahn et al. (2014) on data repositories in South Africa. The findings underscored the fact that South Africa has established a number of data repositories to manage research data, including the South African National Park, the National Health Information Repository and Data Warehouse and Data Intensive Research Infrastructure for South Africa. Uganda's research institutions have a functional Institutional Repository and digitization unit (Mukiibi, 2016); their policies and guidelines are silent about RDM. It is important that The Uganda National Council of Science and Technology (UNCST) establishes an institutional repository for research datasets with clearly spelled out policies to facilitate RDM in collaborative research work in Uganda.

Agricultural research institutes need to invest aggressively in data repositories in consultation and in collaboration with the government, research institutes and RDM stakeholders in order to change the landscape of scholarly communication across agricultural research institutes.

3.6.6 Data access

Data access entails distributing data, sharing data, publishing data, linking data to outputs, controlling access, establishing copyright, and promoting data (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). Open access refers to the practice of making peer-reviewed scholarly research and literature freely available online to anyone interested in reading it (European Commission, 2016; and Organization for Economic Co-operation and Development (OECD), 2007, 2004).

In open access, agricultural research data is increasingly declared to be a ‘public good’ which is worthy of preservation as long as it represents scientific or socio-economic value; this poses new questions on selection and on who should pay for long-term preservation. In this regard, Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Group (2016) noted that open access has been enhanced by a proliferation of data repositories to store and preserve research data for reuse and sharing at local, national and international levels. Access to research data plays a pivotal role in RDM in any given agricultural research institute. It is imperative that open access movement (Berlin Declaration, 2003; European Commission, 2016; Higman and Pinfied, 2015) have advocated for research data to be carefully preserved and made widely available through open access to enhance sharing, use and reuse. Furthermore Joint (2008) underscores the role played by open access repositories in promoting preservation, access and use of agricultural research data.

European Commission (2016) identified the open access mandates to comprise two steps:

a) Depositing publications in repositories (researchers must deposit a machine–readable electronic copy of the published version of agricultural research data in a repository for publication);
b) Providing open access to interested persons {researchers must ensure open access to the deposited publication via the chosen repository either through self-archiving (green open access) or open access publishing (gold open access)}.

A case study of five disciplines by RECODE Project Consortium (2014) found that the development of open access to research data needs to be informed by research practices and processes in the different disciplines and characterized by collaborative partnership by key stakeholders and other research institutes (RECODE Project Consortium, 2014). This will ensure participation by a wide range of research communities and the embedding of open access within research practices and process. The Scholarly Publishing and Academic Resources Coalition (n.d) lists three key requirements for open access; first- availability (research data should be made available), second- access (research data should be available in a form that is convenient and modifiable) and third- redistribution and reuse (research data must be provided under terms that permit full redistribution and reuse).

In contrast, RECODE Project Consortium (2014) found out that most research institutions in the European Union focus on open access to publications rather than research data and therefore there is need to develop policies that allow the openness of research data but safeguarding the intellectual property rights, ownership while meeting the funders requirements. A review of the literature brings to the fore the importance of open access policies with regard to establishing those authorized to view, edit, download, upload or reuse the research data. Therefore, agricultural research institutes need to implement authority control procedures which allow only those with relevant permission to access research data and guard against illegitimate access and use. Similarly, Karolis (2013) underscores the effects of open research data initiatives, with the potentiality to contribute to more effective open data policies and targeted data reuse efforts.

In support of the principle regarding open scientific research data, the Agricultural Information Management Standards (AIMS), (2013) states:

> …to the greatest extent and with the fewest constraints possible publicly funded agricultural research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.

There are also institutional and community benefits provided by open data in terms of costs, greater accessibility and long-term preservation of research output (Macdonald and Martinez-uribe, 2010; Ball, 2012; and NSF, 2011). Such benefits include:

a) Visibility in terms of increased citation and usage, and greater public engagement. Studies have shown that open access content attracts more attention than non-open access content;
b) Make new discoveries through faster impact, wider collaboration and increased interdisciplinary conversation. Open access papers accelerate the pace of scientific enquiry; and

c) Comply with funder mandates. Increasingly, the funders of agricultural research are requiring that their grant holders make articles related to research available to the public, free and without restriction on reuse.

Muinde and Gorman (2009) in a qualitative study examine barriers to open access to agricultural information in Kenya, found that many agricultural researchers were yet to come to terms with e-communication including e-publishing, open access initiatives and software that aid free sharing of research data. The study further established that there is lack of institutional framework and policy guidelines to guide online communication; ICT infrastructure aiding research data communication and open access are not a priority in budget allocation; researchers lack skills to search and manipulate online information systems; no research data repositories can provide a platform for research data and knowledge sharing; and full text open access journals are limited. Mugata (2014) agrees with Muinde and Gorman that while there has been a proliferation of open data portals, data reuse tools and applications of tremendous speed in the last decade; research data and understanding about the impact of opening up and sharing agricultural research data in Kenya has been deficient. Inadequate storage capacity for research data, confidentiality concerns, intellectual property rights and complexity are some of the stumbling blocks to open sharing of data.

In sum, open access is a means to advance knowledge, increase the benefits and return of investment in research and to foster innovation. Open access to agricultural research data from public funding should be easy, user-friendly and preferable internet based (Thanos, 2010), but this will only be possible if policy, legal and technological dimensions are addressed, for instance, technology must render physical and semantic barriers irrelevant, while policies and laws must address and supplant outdated legal jurisdictional boundaries.

3.6.7 Data use and reuse

Data use and reuse entails follow-up research, new research areas, undertaking research reviews, data citation, scrutinizing findings, and use and reuse of research data for teaching and learning (Van den Eynden, 2013; Van Wyk and Van der Walt, 2014). The use of ICTs ensures that data are accessible to designated users for first time use and reuse bearing in mind that some material may be publicly available, whilst other data may be password protected. The use of data collected in addressing emerging issues or being reused to find out whether a research establishes same conclusions or re-analyses of existing data to come up with powerful insights is imperative to RDM and agricultural research institutes (Lewis, 2010; and Heidorn, 2011).

Good RDM is key to data use and reuse. Australian National Data Service (n.d (a)) explains how data are made available for use and reuse through the following ways:

a) Planning for use and reuse and publication from the start;
b) Appropriate recognition of others' data through appropriate citation;
c) Appropriate rules of use and reuse through simple and explicit data licensing approaches;
d) Sufficient metadata describing how the data has been specified, collected, analyzed and transformed;
e) Use of standard vocabularies in the metadata also enables use and reuse; and
f) The most effective way to get data used and reused is to publish it.

Improvement in technology, tools and communications have made agricultural research data easier to use and reuse. Since agricultural research data is data-rich, researchers have the opportunity to research into many aspects, including re-analysis of existing data, verification of results, minimization of duplication of efforts and acceleration of innovation, leading to improved food production and attainment of food security.

3.6.8 Data transformation

Data transformation or Data mining or knowledge discovery is the practice of examining large databases in order to generate new information; it is the re-analysis of data for relationships that have not previously been discovered. Higgins (2012) points out that the synthesized research data inevitably transforms the raw dataset to create a new one. This demonstrates that transformed datasets can perform a variety of roles such as verifying the results obtained from the analysis of the raw data, forming the basis of further experiments, or forming the basis of a new hypothesis. Witt (2008); Higgins (2012); and Palmer et al., (2007) agree that transformation of agricultural research data finds the data curation lifecycle restarted through the creation of new derived datasets and the reuse of data to underpin new research.

Data transformation in RDM provides user-oriented access to new and hidden patterns in data, from which knowledge is, generated which can help with decision making in an agricultural research institute. Milovic and Radojevic (2015) contend that ICTs have made data transformation more efficient and effective through automated extraction of data in an effort to obtain knowledge and trends. Agricultural research data, if analyzed from different perspectives, help find connections and relationships in seemingly unrelated data that will lead to innovative ideas and new discoveries (Maru, 2004; and Ashurst and Mbithi, 2010). Therefore, through data transformation, agricultural research institutes will be able to provide divergent information and knowledge to agricultural stakeholders such as government, agricultural extension offices, and farmers in improving food productivity and a country's food security.

It is clear that making research data broadly accessible and fully available for reuse encourages new research through the reanalysis of existing data, further leveraging the value of a research investment, promote interoperability and allow the data to be mined using cutting-edge computational tools across huge amounts of data to find connections, trends and patterns (Lyon, Patel and Takeda, 2014).

To bridge the gap in literature, the second research question (see section 1.5.1.1) sought to assess how agricultural research institutes capture, appraise, describe, preserve, access and reuse its research data.

3.7 Sharing research data

Sharing research data facilitates new scientific inquiry, avoids duplication in data collection, and provides rich real life resources for education and training because many research data have a significant value beyond usage for the original research (Jahnke and Asher, 2013). Sharing of data may lead to discovery and use outside the discipline in which the data were created furthermore, sharing fosters interdisciplinary research and learning (Witt, 2008). Concurring with Witt (2008), Van den Eynden et al. (2011) buttress the following as benefits of sharing research data:

a) Encourages scientific enquiry and debate;
b) Promotes innovation and potential new data uses;
c) Leads to new collaborations between data users and data creators;
d) Maximizes transparency and accountability;
e) Enables scrutiny of research findings;
f) Encourages the improvement and validation of research methods;
g) Reduces the cost of duplicating data collection;
h) Increases the impact and visibility of research;

i) Promotes the research that created the data and its outcomes;
j) Can provide a direct credit to the researcher as a research output in its own right and
k) Provides important resources for education and training.

There is, therefore a need for agricultural research institutes to provide not only structures and policies for research data sharing, but services to support and educate researchers on concepts of data management and strategies for sharing data that can often be vital for the continuation of research (Karasti, Baker and Hakola, 2006). The ease, with which digital research data can be stored, disseminated and made easily accessible online, means that many institutions should strive to share agricultural research data to enhance the impact and visibility of the research. The majority of the research funders in their research data sharing policy and mandates insist that publicly funded research data should, as far as possible, be openly available and encourage researchers to share data and outputs to the scientific community and stakeholders (American Psychological Association, 2015). In UK, for instance, funding bodies such as the Economic and Social Research Council (ESRC) and Natural Environment Research Council (NERC) have research data policies in place which encourage researchers to share their research data in a timely manner, with as few restrictions as possible (Lewis, 2010).

Van den Eynden et al., (2011) describes various ways of sharing research data including:

a) Depositing them with a specialist data centre, data archive or data bank;
b) Submitting them to a journal to support a publication;
c) Depositing them in an institutional repository;
d) Making them available online via a project or institutional website and
e) Making them available informally between researchers on a peer-to peer basis.

Approaches to data sharing may vary according to research environments and disciplines, due to the varying nature of data types and their characteristics. It is within this context that a group of major international organizations namely FAO, Global Forum for Agricultural Research (GFAR), The Centre for International Cooperation in Agricultural Research for Development (France) (CIRAD), Department for International Development (United Kingdom) (DFID) and Consultative Group for International Agricultural Research (CGIAR) came together to address the issue of coherence and capacity in developing countries' agricultural research system. In January 2008, they came up with a shared identity called Coherence in Information for Agricultural Research for

Development (CIARD). CIARD vision is “to make public domain agricultural research information and knowledge truly accessible to all” with the aim of enabling organizations and individuals that create or possess public agricultural information and knowledge to share it more effectively (Edge et al., 2011).

In the same vein in Kenya, five principal institutions, namely KALRO formerly Kenya Agricultural Research Institute (KARI), Kenya Agricultural Research Institute-National Agricultural Research Laboratories (KARI-NARL), Kenya Forestry Research Institute (KEFRI), Ministry of Agriculture (MoA) and Jomo Kenyatta University of Agriculture and Technology (JKUAT) created a shared open repository of the outputs of agricultural research through a collaborative partnership named the Kenya Agricultural Information Network (KAINet) which was initiated in 2006 (Chisenga et al., 2011). The mission of KAINet is to promote open access and sharing of information as espoused by CIARD initiative. The establishment of KAINet was in response to demand from the national and international community to promote agricultural research data, information exchange and access among stakeholders in the agricultural sector to support decision-making as well as to promote innovation in agriculture. Chisenga et al., (2011) illuminated that KAINet is Kenya’s response to the CIARD global initiative to promote open access and sharing of agricultural research data.

A qualitative study by Steinhart (2011) established that sharing agricultural research data has the potential to facilitate collaborative approaches in conducting research and when data is shared more widely, holds the potential to advance within a given discipline and even across disciplines. In spite of the potential benefits of sharing agricultural research data, there are barriers owing to a number of factors namely: cultural or sociological (sharing may not be the norm in certain disciplines), procedural (confidentiality or commercialization concerns may mitigate sharing), technology (suitable and accessible infrastructure may simply not exist) or logistical (researchers lack the skills and/or time to share data) (Steinhart, 2011).

Barriers to data sharing and preservation are often due to personal belief and views on the process of sharing or withholding data. The above findings by Steinhart (2011) are shared by Enke, Thessen, Kerstin, Bendix, Seegor et al. (2012); Campbell, Clarridg, Gokhale, Birenbaum, Hilgartner et al. (2002) in a study done entitled ‘The User’s view on Data Sharing’ revealed diverse mix of both technological (lack of appropriate databases) and sociological (time, funding) factors that could impede agricultural researchers from sharing data. More importantly, the main reasons for not sharing data as cited by Enke et al., (2012) and Tenopir, Birch and Allard (2012) in their international survey were:

a) Loss of control over the data;
b) Lack of research institute's-wide standards for data sharing;
c) The amount of time that would need to be invested in sharing data sets;
d) Researchers often have major concerns with legal issues, misuse of data and incompatible data types that interfere with the practice of sharing their data;
e) Researchers and other RDM stakeholders may lack the knowledge about handling data management plan, for example, able to construct an articulate data management plan and meeting its requirements from the funding agencies.

3.8 Knowledge, skills and training requirements of RDM

The Digital Curation Centre (DCC) lifecycle model can only exist within an institutional framework that has the personnel with the correct skills, knowledge and training to perform each of the stages in the lifecycle on a regular basis. Most extant studies have focused on the training of librarians, archivists and researchers to manage metadata, research data archives, data repositories, data curation, searching and retrieval, access, and web portals (Kuusniemi, Heino and Larmo, n.d; Patrick, Byatt, Luca, White, Rumsey et al., 2013; Cox and Pinfield, 2014; and Molloy and Snow, 2012). Moreover, knowledge, skills and training requirements of RDM is informed by CCM framework focusing on knowledge, skill and training required by researchers, RDM stakeholders, librarians and IT specialist in RDM whereas DCC lifecycle model defines roles and responsibilities needed at every stage of data curation lifecycle.

Gold (2007) asks, 'what skills are required for RDM? Is domain expertise essential to effectively work with researchers in RDM? Gold argues that it makes much more sense to train domain experts in curation skills and RDM than it does to try to teach non-domain librarians or archivist to understand the infrastructure and service needs of a domain. Besides, RDM skills such as standards, metadata, discovery services, preservation, searching and retrieval, and service provision are accepted as vital, underpinning the success of RDM. The Society of College, National and University Libraries (SCONUL) (2015) advance the views of the Digital Curation

Centre (DCC) and Research Libraries in United Kingdom on the knowledge and skills required in supporting RDM. They highlighted the following knowledge and skills required of RDM staff:

1. *RDM knowledge:*

a) The research data cycle and the stages of a research career in order to understand where support might be most needed;
b) RDM principles, including RDM planning, curation and preservation;
c) The discipline-specific nature of data; and
d) Open access and data sharing.

2. *RDM skill:*

a) Advocacy, negotiation and diplomacy to work with researchers and other professional staff;
b) Guidance and training to support researchers in carrying out their responsibilities;
c) Advice and guidance to assist with necessary processes and procedures;
d) Understanding researcher requirements in order to anticipate and provide appropriate support;
e) Bibliometrics; and
f) Monitoring data reuse, citations and impact.

Top on the above RDM knowledge and skills, SCONUL (2015), suggests a focus on RDM duties and responsibilities that include:

(a) Collaborative strategies planning
(b) Policy development
(c) Project management
(d) Auditing good practice and/or provision of services;
(e) Identifying demand for data storage, analyzing requirements for RDM services;
(f) Developing, managing, and maintaining a data repository;
(g) Developing procedures and work flows for RDM;
(h) Designing, developing and providing research training;
(i) Providing advice and guidance for researchers
(j) Liaison, networking and communication
(k) Advocacy with researchers, other professional services and key stakeholders; and

(l) Building communities of practice.

The RDM duties and responsibilities could assist the agricultural research institute generate the knowledge, skill and training required to effectively execute RDM.

Patrick et al., (2013) conducted a survey of thirty seven (37) RDM staff on training to support researchers on RDM at the University of Oxford and University of Southampton and concluded that there is need for RDM staff to have additional training on:- planning for RDM (use of file naming to assist file management and retrieval, compliance with RDM policy, version control of files so that it forms good practice and more); RDM during the research process (identifying the data storage requirements, awareness on data security issues, and description of data for retrieval purposes); and post-RDM (options for the dissemination and sharing of data, licensing of the data for reuse, long-term preservation and archiving of data. Just as important, RDM staff involved IT services, libraries, archives and research services needs to ensure that adequate advice and assistance is available to researchers such that there are no gaps in the support.

According to a study done by Henty (2014) on RDM competencies, the training needs of RDM occasioned by the influence of ICTs needs to address data generation, processing, preservation, dissemination, sharing and reuse. This view is consistent with the argument advanced by Taylor (2014) that skills acquired to curate, access, reuse, and share research data with the advent of new technologies remains one of the major challenges. Taylor continues to explain that researchers, librarians, archivist and IT specialists are deficient in knowledge and skill that pertains manipulation, interpretation and long term access to research data collected. The right skills still need to be recruited and developed, and this would only be possible if the RDM stakeholders have a good understanding of the niche areas that need to be occupied.

Researchers often require training to enable them to acquire knowledge and skills needed to make their research data available and accessible or how to reuse data and incorporate data in their research process. Also, librarians, IT specialists and archivists require training on provision of research data services (RECODE Project Consortium, 2014). Therefore, agricultural research institutes should engage in professional development activities for RDM through workshops, seminars, as well as formal training programs and curricula that enable the gradual development of research data-scientists, RDM skills and knowledge.

The situation in Africa in general and South Africa in particular is a little bit different. Studies by Kahn et al. (2014); Pienaar (2010); and Lötter (2014) on South African exploratory awareness exercise on RDM revealed that awareness of RDM is relatively good but knowledge and skills on policy and resources are wanting. Lyon (2012) and Lewis (2010) advances the view that innovative approaches are needed to address the significant skills gaps, data literacy and training in areas of data curation, open access repositories, research data services and reuse of research data. In addition, Kahn et al., (2014) concurs with Lyon and Lewis by suggesting that skills, knowledge and resource gaps can be addressed by training and development through training providers who have the expertise and capacity to highlight the need for development of relevant curricula and training programs or inclusion of the skill sets specific to RDM into existing programs.

Mukiibi (2016) acknowledges that curricula in most Ugandan universities offering Computing, Information Science lack RDM studies. Mukiibi (2016) also noted that some curricula have incorporated aspects of ICT, librarianship, and archives and record management but silent about RDM studies at any academic certification level. This is evidence of human resource inadequacies across the country, especially with respect to relevant RDM skills and knowledge. These sentiments are shared by Cox, Verbaan and Sen (2012b) who observed that new graduates entering the profession require grounding in RDM related knowledge and skills, but there is also a need requiring professionals to update their competencies.

The gradual development of agricultural research data-scientists, the provision of training and support for RDM is a major challenge though it is necessary to have a combination of regular courses integrated into the curriculum and training courses for researchers, librarians, archivists, and IT specialists.

3.8.1 RDM advocacy, awareness and data literacy

As governments, funders, and agricultural institutes become more involved in RDM, it would be necessary for researchers, librarians, IT specialists and ethics committees to be made aware of the potential benefits of RDM and RDM processes and requirements. Moreover, Kahn et al., (2014) posit that awareness and capacity-building in South Africa through seminars, workshops and conferences promotes the growth of knowledge in this area.

Flores et al., (2015) observed that advocacy is a multidirectional process that learns about researchers needs and conveys the same to the management of agricultural research institutes, while communicating the RDM policy back to researchers and research support units. Consequently, the library should take a leading role in RDM development through advocacy, awareness and training of other RDM stakeholders in agricultural research institutes, especially with regard to utilizing research data generated through access, use and reuse. Advocacy for open data sharing, contributes to the management of an institutional data repository, data literacy among others issues that librarians and archivist should take into consideration (Cox, Verbaan and Sen, 2012b).

Agricultural research institutes should strive for RDM advocacy which is needed at many levels of management, researchers and other stakeholders to share experiences on RDM as well as discuss current and future challenges related to data curation, sharing, open access, data repositories to name a few. On the other hand, data literacy is seen as a new sub-discipline within RDM that emerges from the need to educate researchers and other RDM stakeholders on data curation, sharing and reuse. Data literacy encompasses orientation of researchers, librarians, IT specialist and RDM staff to practically all activities related to RDM, including data curation, data citation and fostering of data quality (Koltay and University, 2015).

To bridge the gap in literature, the third research question (see section 1.5.1.1) sought to examine the knowledge, skills and training requirements for capturing, appraising, describing, preserving, accessing and re-using data.

3.9 ICT infrastructure for RDM

The role of ICT in RDM is an enabler in the management of research outputs. Studies done by Jetten, (2014); Witt, (2008); European Strategy Forum on Research Infrastructure (2009); Amorim et al. (2015); Fary and Owen (2013) and Qin (2013) have consequently focused on ICT infrastructure such as web portal, data repositories, networking, metadata, software and hardware for RDM. These studies used qualitative and quantitative epistemologies and underlined the importance of ICT infrastructure and its compatibilities as critical factors for RDM in agricultural research institutes. Meanwhile, ICT infrastructure for RDM is underpinned by CCM framework by focusing on technical infrastructure that facilitates RDM.

The digital revolution is transforming the way scientific research is conducted. Henty (2014) illuminated that the growing contribution of Information and Communication Technology(ies) (ICT(s)) to research has excited researchers the world over as they invest in new ways of conducting research and enjoy the benefits of more sophisticated computers and communications systems that support measurement, analysis, modeling, simulation, collaboration and publishing. What's more, government and agricultural research institutes are keen to support research in order to ensure that research data are well managed, readily accessible and available on open access by investing in ICTs. More so, if the government is paying for the research, then the public should be entitled to have access to all the products of research.

It is important in today's agricultural research environment for researchers to embrace the use of ICTs in order to effectively collect, analyze, preserve, share and manage research data while at the same time being able to access multi-scale, multi-discipline and multi-national research data. Australian government (n.d) asserts that development in ICTs is revolutionizing the agricultural research sector by setting up high-speed networks, web portals, metadata, and data repositories bringing in substantial potential benefits in data generation, analysis, manipulation, sharing and reuse. A qualitative study by Piennar (2010) conducted at the University of Pretoria on RDM revealed that relevant ICT infrastructure capability and institutional frameworks are important resources in RDM. Amorim et al. (2015) emphasize the need for compatibility of data repositories, metadata, security systems, data management systems, search mechanisms and community acceptance as they are central to RDM.

The advance of digital technologies has both strengthened the power and reach of agricultural research data. A study by Tenopir, Birch and Allard (2012) on research data services on Association of College and Research Libraries (ACRL) in United States (US) and Canada adapting qualitative and quantitative epistemologies, found out that research institutions face many challenges while attempting to preserve the vast amount of data for long-term use, including how best to describe data in a consistent way, keep up with evolving data standards, consistently and effectively share data while allowing some restrictions, obstacles to data sharing and data reuse, all while coping with the huge increases in the amount of data being created. In support of Tenopir, Birch and Allard (2012), Hentry (2014) in his study on ICT infrastructure observed that RDM have not always been matched by corresponding ICT infrastructure support at institutional levels. Similarly, the level of capability among those with support responsibilities has not been able to keep up thus falling short of ensuring that data is managed responsibly and sustained for later discovery and use.

A survey conducted by Maru (2004) on ICT in agricultural research and development in Sub-Saharan Africa found out that National Agricultural Research System (NARS) have major gaps and weaknesses in ICT infrastructure, such as

a) Capability, including skills and training;
b) Content, including generation and management;
c) Capital, with focus on funding;
d) Connectivity, not only physical but the ability to access information buy individual and user community;
e) Organizational or institutional culture; and
f) Conceptual framework related to ICT.

Furthermore, CIARD (2012) did a case study on KAINet and established that the use of Web 2.0, You Tube and Social Networking to enhance visibility and exchange of research outputs, including metadata, has not been widely implemented and embraced by most of Kenya's agricultural research institutes. A major function of the NARS is to effectively use ICT because modern agriculture is knowledge intensive through acquisition, storage, processes and management of information. In Kenya, the lack of appropriate ICT policy and infrastructure for ICT in NARS has been identified as one of the main causes of low technology uptake and the broadening gap between knowledge and application (Kenya, Republic of: Agricultural Sector Coordination Unit (ASCU), 2012).

3.9.1 Research data security

The protection of agricultural research data is a fundamental responsibility, rooted in regulatory and ethical principles and should be upheld by all data stewards. Moreover, agricultural research data security guidelines should help researchers and other RDM stakeholders understand the sensitivity of the research data and develop appropriate data protection plans, know appropriate mediums and places to store data, prepare their research data for public use, and what to do in the event of theft, loss, or unauthorized use of confidential agricultural research data (Princeton University, 2017).

Physical security, network security and security of computer system and files all need to be considered to ensure security of data and prevent unauthorized access, changes to data, disclosure or destruction of data. Data security arrangements need to be proportionate to the nature of the data and the risks involved. Data security may be needed to protect intellectual property rights, commercial interests, or to keep personal or sensitive information safe. Van den Eynden et al. (2011); and Princeton University (2017) explains ways of securing data:

(a) Physical data security
- Controlling access to rooms and buildings where data computers or media are held; and
- Logging the removal of, and access to, media or hardcopy material in store rooms.

(b) Network security
- Not storing confidential data such as those containing personal information on servers or computers connected to an external network, particularly servers that host internet services; and
- Firewall protection and security-related upgrades and patches to operating systems to avoid viruses and malicious code.

(c) Security of computer systems and files may include:
- Locking computer systems with a password and installing a firewall system;
- Implementing password protection of, and controlled access to data file e.g. read only, administrator-only permission;
- Controlling access to restricted materials with encryption; and
- Not sending personal or confidential data via email or through File Transfer Protocol (FTP), but rather transmit as encrypted data.

Researchers and all stakeholders in Kenya's agricultural research institutes working with confidential or restricted data must comply with institute's security policies and protocols to ensure sensitive data is protected. Patel (2016) asserts that research datasets stored and archived need to be protected against hacking, tampering and unauthorized/accidental deletion of research data and on the same note organization should have a well-defined and documented policy for research data security.

More stringent security measures may be applied to data that is confidential or form the basis of a patent application whereas contractual agreements may regulate what disclosures can and cannot be made in relation to research data (Fitzgerald, Pappalardo and Austin, 2008).

3.9.2 Use of ICTs to improve linkages among agricultural researchers, extension workers and farmers

Mugabe (2001) points out that ICTs are creating faster ways of acquiring, storing and disseminating information, thus breaking barriers to knowledge and integration into global economy. According to Munyua (2000) weak linkages between researchers, extension workers and farmers have been a major constraint that has resulted in research findings not being applied by farmers. This view is consistent with the argument advanced by Kassam and Odame (2002) that the full potential of agricultural research is not being realized because communication between scientists, extension workers and farmers throughout developing countries is weak. FAO (2002) shared similar sentiments when it observed that weak linkages between extension and research often resulted in systematic 'knowledge and information bottlenecks' and limit the effectiveness of research to contribute to agricultural development. In this regard, Munyua (2000) asserts that ICTs can improve and strengthen linkages and ensure agricultural research data, knowledge and information, are communicated to all stakeholders.

National Science Foundation (NSF) (2007a) opined that ICT infrastructure capability is ubiquitous, reliable and a widely shared resource operating on national and transnational scales capable of availing research data to farmers via the extension workers. This would facilitate the development of global ICT infrastructure capability that would reduce geographic, socio and national barriers and consequently, facilitate discovery, access and reuse of research data along with enabling researchers to make the best use of the world's growing wealth of data to improve agricultural products.

To bridge the gap in literature, the fourth research question (see section 1.5.1.1) sought to assess the level of ICT preparedness in research data capture, appraisal, description, preservation, access and reuse.

3.10 Collaborative partnerships influencing RDM

Collaborative partnerships are important catalysts in research projects. Grebmer and Spielman (2004); Pinfield, Cox and Smith (2014); Erway and Rinehart (2016); Humphrey (2014); and Flores et al., (2015) state that collaboration within institution and among institutions is necessary for the sharing of research data, and for creating and sustaining public-private partnerships among research institutes, and partners. Collaborative partnerships influencing RDM is informed by CCM framework by focusing on collaboration within the discipline, across disciplines, across sectors and with the public.

Kahn et al., (2014) assert that good RDM may help to make research collaboration between institutions, groups and individuals more efficient by reducing duplication of effort and avoiding data loss. Kahn et al. (2014) continues to affirm that in collaborative research by international agencies and universities in the west have enriched research data repositories in their respective countries as per requirements of their funding agencies. This has long term effects on the agricultural research data sources as after the stipulated time frame that data is lost in host institutions but remains active and available to third parties many years after the initial research. Such agricultural research data can then be retrieved and reused, resulting in new meaning and consequently new findings which gives value to RDM and its accruing benefits (Mukiibi, 2016)

International agencies, for example, the United States (US) National Science Foundation (NSF), Australian National Data Services (ANDS) and the e-Infrastructure Reflection Group in the European Union increasingly recognize that research data, being a pervasive and potentially long-lived information asset for all of society, needs planning, coordination and collaboration with the concerned partners (Australian Government, n.d). In addition, Bracke (2011) illuminates that in the long run, however, RDM potential will only be tapped if the many actors in data creation, management and use are able to develop collaborations to build shared infrastructure and to develop and implement best RDM. Moreover, agricultural research institutes should uphold good governance that would promote collaboration within and across research areas, nationally and internationally, and ensure the effective establishment, operation and management of research data infrastructure.

Pinfield, Cox and Smith (2014) interviewed 26 respondents with regard to collaborative partnerships. The findings revealed that collaborative partnerships on one hand have benefits such as metadata exchange, sharing, and reuse of research data and on the other hand, challenges such as lack of teamwork and policies governing collaborative partnerships. Humphrey (2014) stated that research institutes must ensure that they build collaborative partnerships with other research institutes and funders in order to develop RDM capabilities. Such partnerships use best practices from which to draw lessons and experiences since public-private partnerships tend to be significantly constrained by insufficient accounting of the actual and hidden costs and undue competition over financial and intellectual resources.

RDM is essential and this is particularly the case with collaborative research networks if maximum benefit is to be achieved. In this regard, collaborative research networks in agricultural research institutes create a significant amount of new data and it is imperative that this data is well managed to ensure that it is secure, discoverable, accessible, useable and re-usable (Gibson and Gross, 2013). In light of this, Gibson and Gross (2013) conducted a qualitative study at Edith Cowan University-Australia on the research data challenges and possible solutions within a collaborative research network. The findings revealed the following challenges:

a) Data sharing: the main challenge for research community is sharing research data during collaborative projects among research institutes, inter-state and international partners. Different layers of complexity in sharing research data with collaborators include:
    i. Data containing confidential information;
    ii. The size of data to be shared-variations in terms of the size of data shared ;
    iii. The number and location of partners in the collaboration-the more partners collaborating in a project, the greater the risk of the dataset being contaminated.
b) Data storage: collaborative partners may differ on retention periods, the destruction of research data and the availability of data for reuse. Another challenge is the security and storage of datasets, because of the nature of research and potential for data errors resulting from multiple researchers accessing the data resulting in different copies of the data being used and datasets being modified.
c) Education: Within collaboration, partners may have different views on the best practices for RDM, standard data management plan and different training on collection, organization and effective use of metadata, thus posing a challenge on collaborative partners.

Gibson and Gross (2013) suggests solutions on the above mentioned challenges which include:

a) Developing and implementing RDM system that assists researcher community with data and sharing; and
b) Develop a common research portal that could allow all collaborators secure access.

A survey conducted by Maru (2004) on ICT in agricultural research and development in Sub-Saharan Africa revealed relatively little collaboration and partnerships between and across NARS and between and across the implementing organizations like the government, funders and other international agricultural research institutions. Concurring with Maru, a study by Mugwisi (2013) on Information needs of agricultural researchers and extension workers in Zimbabwe, postulates that there is a wide gap on collaborations between agricultural researchers and extension workers, the public and private, the national and international agricultural research institutes. However, collaborative partnerships in agricultural research institutes is essential in planning future research strategies that are global in coverage and requires cooperation by all entities in agricultural research in order to ensure that limited global resources in agricultural research are used and more so able to collaborate on issues of RDM in order to share agricultural research data to strategically address the issue of food security and agricultural production.

Australian Government (n.d) calls for a national collaborative approach to investment in research data infrastructure in order to reduce duplication, enhance economic use of resources, and optimize research outcomes and benefits. In addition, appropriate access arrangements and agreed standards will facilitate collaboration, fostering multi-disciplinary research uses for existing data, enabling researchers to address emerging problems in new ways. Kenya, Republic of: ASCU (2012) points out that collaborative partnerships influence RDM in agricultural research institutes in Kenya to thrive and be successful while ongoing dialogue between collaborating partners (both internal and external) to ensure needs are understood, documented and acted upon for common benefit. In sum, a joined-up (collaborative partnerships) research data environment will be a significant component of a strong, cohesive research fabric that will support RDM across a broad range of disciplines including the development of research data infrastructure that enables management, sharing and reuse of agricultural research output.

To bridge the gap in literature, the fifth research question (see section 1.5.1.1) in this study sought to examine how collaborative partnerships influence research data capture, appraisal, description, preservation, access and reuse of research data.

3.11 Data management plan

A data management plan is a planning document which is supposed to describe how research data should be collected, stored during research phase, analyzed, described, archived on a long-term basis, access, shared to the public, secured, determine quality assurance and quality control measures (Stanford Libraries, n.d(b)). More importantly, research funders place increasing importance on DPM as a mechanism for improving the longevity of research data and for enabling its widespread access and reuse. Fitzgerald, Pappalardo and Austin (2008) point out that a DMP will need to set out the different security measures relevant to the different levels of data and how this security measures are to be implemented simultaneously addressing the legal and regulatory controls applying to the data generated by the research project. DMP provides many benefits to researchers (Karick, 2014) which include:

a) Decreased risk of data loss or misuse;
b) Good research practice which ensures integrity and quality of data;
c) Enable researchers to identify researcher storage needs;
d) Enables researchers to easily defend their research method;
(a) Enable researchers to be more proactive about their research needs; among others

According to Donnelly (2012), every DMP is different since they reflect the different research project undertaken and research funders. However, the following general guidelines are provided by Donnelly (2012); Karick (2014); Van den Eynden et al. (2011) when developing DMP:

a) Which research data is generated during research;
b) Metadata, standards and quality assurance measures;
c) Plans for sharing data;
d) Ethical and legal issues or restrictions on data sharing;
e) Ownership, copyright, Intellectual Property Right (IPR) of research data;
f) Data storage, backup, security;
g) How to preserve, share and access;
h) RDM roles and responsibilities and
i) Costing or resources needed.

RDM is growing in importance, as a result of funders' requirements of DMP increase in a data sharing culture among researchers, and an ongoing policy shift that requires open access to research both on a national and international level.

3.12 Support services for RDM (library and IT department)

Libraries and IT departments have clear roles in the data science arena especially concerning awareness of data issues and the importance of good data science and curation (Swan and Brown, 2008). Libraries and IT departments supporting agricultural research institutes should clearly outline their mandates with regard to RDM in terms of providing a well-structured and synchronized research data services. According to Tenopir, Birch and Allard (2012), research data services are services that libraries and IT departments offer to researchers and other RDM stakeholders in relation to managing data and providing technical services, such as:

a) Consulting on metadata standards and data management plans;
b) Providing reference support for finding and citing data sets;
c) Proving technical support for data repositories;
d) Preparing data sets for a repositories;
e) Deselecting data sets from a repositories and
f) Creating metadata for data sets.

Several studies have cited the importance of library staff training in the area of data curation and management services. For example, Newton, Miller and Bracke (2011) in their exploration of the librarian's role in RDM, the study found strong evidence that librarians and IT specialists needed additional skills in data competency areas, identifying and collecting data and data sets to include in repositories, technical hands-on training in the digital description and curation of large data sets. The rapid changes in the agricultural research landscape make continued research data services offered to researchers a necessity. A comprehensive and strategic role for libraries and IT departments should provide for active participation in data curation, promoting access, sharing and reuse (Tenopir, Allard, Douglas, Aydinoglu, Read et al., 2011). Therefore, research data services should address the full data lifecycle with reference to DCC lifecycle Model (Higgins, 2008).

Developing and implementing RDM policy and developing relevant services is a team effort that requires the collaboration of multiple actors including researchers, IT specialists, librarians, agricultural research institute leadership and other RDM stakeholders. In this regard, RECODE Project Consortium (2014) recommended that libraries and IT departments should be involved in operationalizing policies, developing technical infrastructure and services, training researchers, offering awareness and advocacy services to promote research data curation, access, sharing, reuse and disposal. This new environment allows librarians and IT specialists to take a more active and visible role in the knowledge creation process by placing them at all stages in the research planning process and providing expertise in identifying appropriate data description, data repositories, data curation, creating preservation strategies, data sharing and reuse in agricultural research institute.

Yakel, Ixchel, Faniel and Yoon (2013) and Flore et al. (2015) affirms that libraries and IT departments are a vocal and critical part of RDM discussion because they are part of the policymaking body and takes on the role of RDM advocate by helping to complete the RDM feedback loop among stakeholders, and offering RDM training and services to researchers.

Moreover, libraries and IT departments in agricultural research institute give support in the scholarly and communication process for researchers, such as publishing advice, preparing for access to researchers' publications, helping researchers, measuring their scientific output, among others. (Lyon, 2012).

3.13 Scholarly communication

According to the Association of Research Libraries (2014) Scholarly communication is defined as the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use. Warden (2010) refers to scholarly communication as the process by which academics, scholars and researchers share and publish their research findings so that they are available to the wider academic community and beyond.

Western Libraries (2013) point out that there are different stakeholders involved in the modern scholarly communication process including authors, publishers, libraries, researchers, research institution, and funding agencies. In the context of this study, Scholarly content include agricultural research data which may be in the form of text, sound, still images, moving images, models, simulations, databases that require computational machinery, observational data and socioeconomic data (Borgman, 2012). RDM must support scholarly communication (Lord and Macdonald, 2003) as it is composed of the following:

a) The original researcher produces, through research activity, primary raw data;
b) Data is analyzed to create secondary data;
c) This is then evaluated, refined, to be reported as tertiary information for publication;
d) With the mediation of the pre-print and peer review mechanisms, then sets out into the traditional publishing process and feeds publication archives.

With the global proliferation of scholarly content from agricultural research institutes, it is clear that strategies for its management need to be carefully considered to enable the communities in and outside these agricultural research institutes to benefit from the knowledge generated. Van den

Eynden and Brett (2010) posit that scholarly community benefits from research data sharing in the following ways: valuable resources in teaching; promote innovation (reuse of data); quality improvement from verification, replication and trustworthiness; and maintain professional standards of open inquiry. Besides, Loughborough University (2017) points out that RDM contribute to scholarly communications through the following ways: increase research visibility, increase citation count, increase research efficiency, leads to co-authorships, and builds collaborations. Therefore, researchers need to be fully cognizant of the benefits of RDM in order to re-assert themselves in the scholarly communication process.

Summary

The chapter has reviewed the various empirical and theoretical literatures pertinent to the subject under the study. The literature reviewed revealed that RDM have been recognized worldwide, albeit to different extents. Moreover, the literature reviewed established that globally, the research institutions are under pressure to embrace RDM to facilitate research data capture, appraisal, preservation, access, sharing and reuse. Additionally, research funders place increasing importance on RDM as a mechanism for improving longevity of research and widespread access, sharing and reuse.

Moreover, literature reviewed revealed that for the successful establishment of RDM in agricultural research institutes, enactment of RDM legal, policies and regulations should be in place, adoption of ICTs to fast track RDM, collaborative partnerships in boosting relationships among research institutes and knowledge, skill and training are the core muscles behind the success of RDM. A review of the literature brings to the fore the following gaps that exist in different themes: further investigations on legal, policies and regulation framework governing RDM; further analysis on data curation, sharing and reuse; the need to re-look and examine knowledge, skills and training for RDM; further studies on ICT infrastructure for RDM; and more analysis on how collaborative partnerships influence RDM. However, in order to bridge the gaps in literature, the study sets out research questions that addresses these gaps. The research questions were further subjected to an empirical study whereby in-depth investigations and analysis on the gaps were under taken, the findings of which are discussed in chapter 5 and 6. Therefore, literature review contributed enormously to the success of this study by according the researcher a thorough understanding of RDM in agricultural research institutes.

The next chapter provides detailed discussion on research methodology of the study.

CHAPTER FOUR

RESEARCH METHODOLOGY

4.1 Introduction

Industrial Research Institute (2010) defines research methodology as a way to find out the result of a given problem on a specific matter or research problem. Similarly Rajasekar, Philominathan and Chinnathambi (2006); and Kothari (2004) refer to research methodology as the procedure by which researchers go about their work of describing, explaining and predicting phenomena. In this regard research methodology aims at studying the various steps that are generally adopted by a researcher in studying his/her research problem along with the logic behind them.

The purpose of this study is to examine research data management practices in Kenya's agricultural research institutes with the view to proposing interventions to improve management, sharing and reuse of agricultural research output. This chapter is organized into the following thematic sections: research paradigm, research methods, research design, population of study, sampling procedures, data collection techniques, data collection procedures, data analysis strategies, validity and reliability of data collection instruments, ethical consideration and summary.

4.2 Research paradigm

All research is based on some underlying philosophical assumptions about what constitutes valid research and which research methods(s) is/are appropriate for the development of knowledge in a given field of study. In order to conduct and evaluate any research, it is therefore important to know what these philosophical assumptions/paradigms are. Bryman, Teevan and Bell (2009) define research paradigm as a set of beliefs and assumptions about how the world works and how knowledge of it is to be gained. According to Rossman and Rollis (2003), the term paradigm refers to a worldview or a set of assumptions about how things work or shared understandings of reality. Patel (2015) views a research paradigm as the set of common beliefs and agreements shared between scientists about how problems should be understood and addressed.

Kuhn (1970) advance the view that paradigm refers to a research culture with a set of beliefs, values and assumptions that a community of researchers has in common regarding the nature and conduct of research. A paradigm hence implies a pattern, structure and framework or system of scientific and academic ideas, values and assumptions (Olsen, Lodwick and Dunlop, 1992).

Weaver and Olson (2006) sum it up by stating that research paradigm are patterns of beliefs and practices that regulate inquiry within a discipline by providing lenses, frames and processes through which investigation is accomplished.

Guba (1990) characterized research paradigm to comprise the following components in Figure 4.1:

Ontology - What is reality?

Epistemology – What and how can I know reality/knowledge?

Theoretical perspective – What approach can we use to get knowledge?

Methodology – What procedure can we use to acquire knowledge?

Methods – What tools can we use to acquire knowledge?

Source – What data can we collect?

Figure 4.1: Components of research paradigm (Source: Guba, 1990)

Guba (1990) asserts that ontology and epistemology create a holistic view of how knowledge is viewed and the methodological strategies used to un/discover it. Patel (2015); Morgan (1998) and Creswell (2003) in this regard classified research paradigms into four philosophically distinct categories namely:

1) Positivists paradigm believe that there is a single reality, which can be measured and known, and therefore they are more likely to use quantitative methods to measure. In essence positivist has the elements of being reductionistic, logical, empirical, cause-and-effect oriented, and deterministic.
2) Social constructivist paradigm (interpretivism) holds the assumption that individuals seek understanding of the world in which they live and work and relies as much as possible on the participants' views of the situation being studied.
3) Pragmatist paradigm believes that reality is constantly renegotiated, debated, interpreted, and therefore the best method to use is the one that solves problem. In essence it focuses

attention on the research problem then uses pluralistic approaches to derive knowledge about the problem.

4) The advocacy and participatory paradigm deals with marginalized individuals in society/issues of social justice that need to be addressed. The research should contain an action agenda for reform that may change the lives of the participants or institutions in which individuals work or live.

The current study is underpinned by the pragmatism paradigm as motivated in section 4.2.1.

4.2.1 Pragmatism paradigm

Pragmatic paradigm recognizes that there is a concern with application-'what works'-and solutions to problems. Pragmatic paradigm therefore focuses on the problem being studied and the questions asked about the problem (Patton, 1990; and Rossman and Wilson, 1985). In discussing pragmatism, Cherryholmes (1992); Margolis (2003); and Murphy (1990) provide direction for the basic ideas:

a) Pragmatism is not committed to any one system of philosophy and reality: The inquirers draw liberally from both quantitative and qualitative assumptions when they engage in research;
b) Individual researchers have a freedom of choice. They are free to choose the methods, techniques, and procedures of research that best meet their needs and purposes;
c) Pragmatists do not see the world as an absolute unity instead, researchers look to many approaches to collecting and analyzing data rather than subscribing to only one way.

From this point of view, the study adopted pragmatism paradigm because it focuses attention on the research problem and the use of pluralistic approaches to derive knowledge about the problem (Rossman and Wilson, 1985; and Creswell, 2009). The study sought to examine Research Data Management (RDM) in Kenya's agricultural research institute with the view to proposing interventions to improve management, sharing and reuse of agricultural research output.

Pragmatism paradigm uses both quantitative and qualitative approaches to focus on 'what' and 'how' aspects of research based on its intended consequences. Creswell (2013) acknowledges that researchers using this worldview (pragmatism) apply multiple methods of data collection to best answer the research question, employ multiple sources of data collection; they focus on the practical implications of the research, and emphasize the importance of conducting research that best address the research problem. The Pragmatic paradigm has been used in similar and related studies. For example Kiplang'at (2004) used pragmatic paradigm to investigate diffusion of Information and Communication Technologies (ICTs) to disseminate agricultural information among agricultural researchers and extension workers in Kenya. Similarly, Maseh (2015) applied pragmatic paradigm to study 'Records management readiness for open government in the Kenyan Judiciary'. The essence of pragmatism ontology is that it focuses on action, change, situations and consequences about the research problem.

4.3 Research methods

Research methods are the various procedures, schemes and algorithms used in research (Rajasekar, Philominathan and Chinnathambi, 2006). Research methods are essentially planned, scientific and value-neutral; they include theoretical procedures, experimental studies, numerical schemes, and statistical approaches. According to Bhattacherjee (2012); Creswell, (2003); and Edmond and Kennedy, (2013), research methods can be categorized into three, quantitative, qualitative methods and mixed method.

Creswell (2003) explains further the three methods outlined above as follows:

a) Quantitative method is one in which the researcher uses post positivist claims to develop knowledge, employ strategies of inquiry such as experiments and surveys and collects data on predetermined instrument that yield statistical data;
b) Qualitative method on the other hand is one in which the researcher makes knowledge claims based on constructivist perspectives. The researcher also employs strategies of inquiry such as narratives, phenomenologies, ethnographies, grounded theory or case studies. The researcher collects open-ended data with the intent of developing themes from the data;
c) Mixed method is one in which the researcher tends to base knowledge claims on pragmatic grounds. It employs strategies of inquiry that involve collecting data either concurrently or sequentially to best understand the research problem. The data collection involves both quantitative and qualitative information.

The current study embraced mixed method where both qualitative and quantitative methods were equally utilized in order to comprehensively collect data that could answer the research problem.

4.3.1 Mixed method

Mixed Method is both a method and methodology for conducting research that involves collecting, analyzing, and integrating quantitative and qualitative research in a single study or a longitudinal program of inquiry (Creswell, 2008). Tashakkori and Teddlie (2010) define mixed method as the broad inquiry logic that guides the selection of specific methods and that is informed by conceptual positions common to mixed methods practitioners. Tashakkori and Teddlie explain further that the definition of methodology distinguishes the mixed method to conducting research from that practiced in either the quantitative or qualitative approach. Topical issues are advanced by Johnson and Turner (2003); and Greene (2007) when they argued that the fundamental principle of mixed method is that multiple kinds of data should be collected with different strategies and methods in ways that reflect complementary strengths and non-overlapping weaknesses, to provide insights not possible when only qualitative or quantitative data are collected.

More importantly, mixed method is the mixing of qualitative and quantitative approaches to provide an all-inclusive analysis of the research problem. Johnson, Onwuegbuzie and Turner (2007) pose the question: What is the reason for using mixed methods? They proffer the following answers:

1. Either quantitative or qualitative may be insufficient by itself
2. Quantitative and qualitative approaches provide different ‘pictures
3. Combined quantitative and qualitative provides more evidence

The study used mixed method to achieve various aims, including corroborating findings, generating more complete data, and using results from one method to enhance insights attained with the complementary method (Creswell and Piano Clark, 2007; and Morgan, 2006). According to Bryman (2006), mixed method uses both qualitative and quantitative research techniques because it was designed to answer both ‘what is it like’ and ‘how many’ type questions. Quantitative analyses employ descriptive and inferential statistics, whereas qualitative analyses produces expressive data that provided descriptive details thus assist in examining the study’s research objectives. According to Creswell and Plano Clark (2011); and Tashakkori and Teddlie (2003), mixed method ensures that data collected through one method can be validated using the other method.

Morse, Wolfe and Niehaus (2006) explain that in mixed method, qualitative and quantitative components may be performed concurrently or sequentially, and emphasis may be placed on either component or equal weight given to both. Hughes (n.d); and Creswell (2003) state that:

1) The *sequential explanatory* method: employs two different data-collection time points with the quantitative data being collected first and the qualitative collected last. Qualitative data are used to enhance, complement and in some cases follow up on unexpected quantitative findings.
2) The *sequential exploratory* design: suited for testing emergent theory because both types of data are interpreted during the data integration phase. This approach is useful when the researcher's interest is enhancing generalisability.
3) The *sequential transformative* approach: There is no preference for sequencing of data collection and emphasizes theory.
4) *Concurrent triangulation:* This method is useful for cross-validation studies with one point of data collection. Qualitative and quantitative data are collected concurrently so that weaknesses of one kind of data can ideally be offset by strengths of the other kind.
5) The *concurrent nested* design: qualitative and quantitative data are collected concurrently and analyzed together. Greater weight is given to one kind of data, in the sense that one kind of data is typically embedded in the other.
6) The *concurrent transformative* approach: is theory driven and allows the researcher to examine phenomena on several different levels. Qualitative and quantitative data are collected concurrently and can be weighted equally or unequally during the integration of findings. Qualitative and quantitative data are typically mixed during the analysis phase.

This study adopted concurrent triangulation. With concurrent triangulation, qualitative and quantitative data were collected concurrently, such that weaknesses of one kind of data were ideally offset by strengths of the other kind. Equal weight was given to the two kinds of data in mixing of the findings. The qualitative and quantitative data were analyzed separately and mixing took place when the findings were interpreted (Hughes, n.d; and Creswell, 2003). Motivation behind the adoption of concurrent triangulation approach was to provide a comprehensive analysis of the research problem by integrating both forms of data in the interpretation of the overall results (Harwell, 2011).

Related studies that have used the mixed method are numerous. For example, Wambani (2011) in a study aimed at examining ways of improving scientific research at KARI-Kakamega research centre adopted mixed with concurrent approach. Moseti (2015) also examined strategies for managing scholarly content at universities in Kenya using a dominant –less dominant mixed method. The quantitative approach was the dominant while the qualitative part was less dominant. Tashakkori and Teddlie (2003) affirm that if mixed method is planned appropriately, each type of data can mirror the other's findings, so the methodology can benefit many types of research.

4.4 Research design

Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design. Kothari (2004) defines research design as the arrangement of conditions for collecting and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.

The function of a research design is to ensure that the evidence obtained enables the researcher to answer the initial question as unambiguously as possible. In other words, when designing research there is need to ask: given the research question, what type of evidence is needed to answer the question in a convincing way (Yin, 1989; and Polit and Beck, 2004). Furthermore, Creswell (2013); and Kothari (2004) explained that there are different research designs which relate to philosophical assumptions and research methods, for instance: research designs associated with Post positivism paradigm and brings into play Quantitative methods are descriptive, correlation, experiment, survey and comparative. Similarly, research design associated with social constructivism paradigm and brings to play qualitative method are ethnographic, phenomenological, grounded theory, case studies and narrative. Finally, the research design associated with pragmatic paradigm which invokes mixed methods is concurrent parallel design, explanatory sequential design, exploratory sequential design, and embedded design (Creswell and Plano Clark, 2011; and Creswell, 2012).

The current study adopted concurrent parallel design as discussed earlier (See 4.3.1) combining the survey design applied within a case study. The mixing of the two designs provided a better understanding of the research problem because it is utilized and built on the strengths of both quantitative and qualitative data (Creswell, 2008; and Saunders, Lewis and Thornhill, 2012).

4.4.1 Case study research design

The case study is a method of study that focuses on in-depth rather than breadth. Leedy and Ormrod (2005) define case study research as a type of qualitative research in which in-depth data are gathered relative to a single individual, program, or event, for the purpose of learning more about an unknown or poorly understood situation. Creswell (2013) defines case study as a type of design in qualitative research that may be an object of study, as well as a product of the inquiry. Creswell further adds that case study research is a qualitative approach in which the investigator explores a real-life, contemporary bounded system (a case) or multiple bounded systems (cases) over time, through detailed, in-depth data collection involving multiple sources of information (interviews, observation, focus groups or document reviews) in which inferences are drawn. Yin (2009) emphasizes that within a case study, the boundaries between the phenomenon being studied and the context within which it is being studied are not always apparent. The case study strategy is relevant if the researcher wishes to gain a rich understanding of the context of the research and the processes enacted (Eisenhardt and Graebner, 2007). Yin (2009); and Henning (2004) assert that case study may use quantitative or qualitative methods and many case study designs use mixed methods to collect and analyze data.

There are four case study strategies (Creswell, 2013; Saunders, Lewis and Thornhill, 2012; and Stake, 1995) namely:

a) Single instrumental case study: the researcher focuses on an issue and then selects one bounded case to illustrate the issue;
b) Multiple case study: the one issue is again selected, but the inquirer selects multiple case studies to illustrate the issue.
c) Intrinsic/holistic case study: the focus is on the case itself because the case presents an unusual or unique situation.
d) Embedded case study: the focus is on the departments or work groups or logical sub-units within the organization whereby the case will inevitably involve more than one unit of analysis.

The current study utilized multiple case study design involving six agricultural research institutes namely:

1) Dairy Research Institute (DRI)

2) Tea Research institute (TRI)
3) Biotechnology Research Institute (BRI)
4) Coffee Research Institute (CRI)
5) Food Crops Research Institute (FCRI)
6) Genetic Resources Research Institute (GeRRI)

Using the case study elements within the agricultural research institutes were comprehensively studied and described (O'Leary, 2004).

Many studies that have widely used case study research design: Jones (2013) in a study of 'the improvement of research data management at Monash University' adopted case study to look at the infrastructure and services in place in order to improve RDM. Similarly, Koopman and De Jager (2016) investigated the archiving South African digital research data using a case study design. Chinyemba (2005) on his part adopted a case study design in a study titled 'managing records at higher education institutions: a case study of the University of KwaZulu-Natal, Pietermaritzburg Campus'.

4.4.2 Survey research design

Survey strategy is one of the most commonly used data-gathering techniques and it is usually associated with a deductive approach. Creswell (2012) defines survey research design as procedures in quantitative research in which investigators administer a survey to a sample or to the entire population of people to describe the attitudes, opinions, behaviors, or characteristics of the population. Saunders, Lewis and Thornhill, (2012) like Creswell posits that the survey strategy allows the researcher to collect quantitative data which can be analysed quantitatively using descriptive and inferential statistics.

There are two types of survey research namely the cross-sectional survey design and the longitudinal survey design. The current study adopted cross-sectional survey design because the researcher collected data at one point in time and measured the attitude or practices at that point in time (Creswell, 2012; and Cohen, Manion and Marrison, 2007). Further survey research typically collects data using two basic forms: questionnaire and interview and in this regard the researcher considered the forms and weighs the advantages and disadvantages of each (Bryman, Teevan and Bell, 2009).

The survey design was applied within a case study in the current study. The choice of a survey design was largely informed by the need to describe the attitudes, opinions and characteristics of researchers regarding RDM in Kenya's agricultural research institutes. The survey design was suitable because it helped to collect data from large number of members of a population (Babbie, 2001). Babbie further points out that survey is a self-report study, which requires the collection of quantifiable information from the sample. Therefore, to collect data from a large number of respondents, survey research design was embraced.

4.5 Population of the study

Population refers to an entire group of individuals, events or objects having common observable characteristics (Mugenda and Mugenda, 1999). In selecting the target population, notions of appropriateness and practicability should be considered (Keya, Makau, Mani and Omari, 1989).

Kenya government has fifteen (15) agricultural research institutions (KALRO, 2016). The population of study comprised directors of institutes, heads of research, researchers, heads of Information Technology (IT) and librarians. Table 4.1 below presents the study population in the respective research institutes.

Table 4.1: The study population in Kenya's agricultural research institutes

Institutes	Directors of Institutes	Heads of research	Researchers	Heads of IT	Librarians	Total
DRI	1	4	40	1	1	47
TRI	1	4	32	1	1	39
BRI	1	4	26	1	1	33
CRI	1	3	43	1	1	49
FCRI	1	4	65	1	1	72
GeRRI	1	4	28	1	1	35
Total	6	23	234	6	6	275

Key: Dairy Research Institute (DRI); Tea Research Institute (TRI)
Biotechnology Research Institute (BRI); Coffee Research Institute (CRI)
Food Crops Research Institute (FCRI); Genetic Resources Research Institute (GeRRI)

(**Source of data**: Human Resources Database, 2015).

The relative distribution of the 275 population was as follows: Directors of research institutes were six (6), twenty three (23) heads of research, two hundred and thirty four (234) researchers, heads of IT six (6) and Librarians six (6).

Directors of research: The directors of research are the persons in charge of the agricultural institute. Responsible for the leadership, support the vision and the mission, implementation of objectives of the agricultural research institute, contribution to public good, enact RDM policies, ensure high-quality research and responsible stewardship.

Heads of research: The heads of research are in charge of research in the research institutes. Their responsibility is to track research output and outcomes, advance and preserve knowledge, promote high-impact research, build reputation and prestige, ensure grant compliance, enhance productivity and increase grant funds, implement data management plan, implement RDM policies, intellectual property administration, oversight of research centers, grants management, and administration of research.

Researchers: Researchers are individuals specialized in a particular subject area/discipline. The researchers ensure grant compliance, creating high-impact research, expanding reach and recognition, advancing and disseminating knowledge, improving research quality and efficiency, accessing data for reuse, verification and review, data sharing, access, and reuse.

Heads of IT: Heads of IT are responsible for IT section/department. Responsible for implementing cyberinfrastructure to support data acquisition, storage, security, integration, preservation, archiving, searching and retrieval, mining, visualization and other information processing service.

Head librarians: Head librarians are responsible Library sections/departments. Librarians are key players in RDM given their experience with selection, metadata, collection, data curation, preservation, institutional repositories, data access, sharing, intellectual privacy, creation of data management plan and trusted stewardship.

4.6 Sampling procedure

Neuman (2000) defines a sample as a selection taken from a large group (the population) so that the researcher can examine it to find out something about the large group. Kothari (2004) defines a plan for obtaining a sample from a given population. Similarly, Mugenda and Mugenda (1999) define sampling as the process of selecting a number of individuals for a study in such a way that the individuals selected represent the large group from which they were selected.

The primary purpose of sampling is to get a representative sample from a much larger population, study it and produce accurate generalizations about the larger group. Becker (1998) posits that in selecting a sample to study, it should represent the full set of cases in a way that is meaningful and which can be justified.

Simple random sampling of the population was adopted in the current study (Saunders, Lewis and Thornhill, 2012) because:

i. It would be impracticable to survey the entire population;
ii. Budget constraints prevented the researcher from surveying the entire population;
iii. Time constraints prevented the researcher from surveying the entire population.

Kenya government has fifteen (15) agricultural research institutions (KALRO, 2016). Saunders, Lewis and Thornhill (2012) point out that sampling provides a valid alternative when it is impractical to survey the entire population due to time and finances. However, the study purposively targeted six (6) Kenya's agricultural research institutes namely:

1) Dairy Research Institute (DRI);
2) Tea Research Institute (TRI);
3) Biotechnology Research Institute (BRI);
4) Coffee Research Institute (CRI);
5) Food Crops Research Institute (FCRI);
6) Genetic Resources Research Institute (GeRRI).

The six (6) Kenya's agricultural research institutes are purposively chosen because their focus is on agricultural research. They have also a long history of undertaking agricultural research in Kenya with different disciplinary areas in agricultures such as livestock, food crops, cash crops, biotechnology, among others. In addition, the researcher selected a manageable sample from the entire population of agricultural institutions due to limited time and financial resources.

Census was used as a sampling strategy for directors of institute, heads of research, heads of IT and librarians. Kapahi (2014) defines census as the procedure of systematically acquiring and recording information about the members of a given population by collecting data from each and every unit. Krishnaswami and Rangnathan, (2010) and Gay and Airasian, (2003) point out that for a small population (with fewer than 100 people or other units) there is little point in sampling and the entire population should be surveyed.

A survey was used to select sample from the researcher's stratum with a population of 234 using Saunders, Lewis and Thornhill (2012) table of selecting sample sizes. Saunders et al. table of selecting sample sizes was found to be more recent. The sample size of researchers was selected based on 95 confidence level and 5% margin for error.
Consequently, the average of 132 and 151 was taken because they lie between the actual 234 and 249:

$$132+151 \qquad = \frac{283}{2} \qquad =142$$

A sample size of 142 respondents was selected for study

Table 4.2: Sample sizes for different sizes of population at 95 confidence level
(Assuming data are collected from all cases in the sample)

	Margin of error			
Population	5%	3%	2%	1%
50	44	48	49	50
100	79	91	96	9
150	108	132	141	148
200	132	168	185	196
250	151	203	226	244
300	168	234	267	291
400	196	291	343	384
500	217	340	414	475
750	254	440	571	696
1 000	278	516	706	906
2 000	322	696	1091	1655
5 000	357	879	1622	3288
10 000	370	964	1936	4899
100 000	383	1056	2345	8762
1 000 000	384	1066	2395	9513
10 000 000	384	1067	2400	9595

(Source: *Saunders, Lewis and Thornhill, 2012:266)*

The distribution of sample sizes in the respective institutions is reflected in Table 4.3 based on the population strength in each institution as illustrated in the given example:

$$DRI = \frac{\text{DRI population x Sample Size}}{\text{Total population}} \text{ (at 5\% error margin)}$$

$$\frac{40 \times 142}{234} = 24 \text{ (DRI sample size)}$$

The rest of the sample sizes per institution are reflected in Table 4.3.

Table 4.3: Relative sample size of researchers in agricultural research institutes

	RESEARCHERS	
INSTITUTES	Population	Sample Size
DRI	40	24
TRI	32	19
BRI	26	16
CRI	43	26
FCRI	65	40
GeRRI	28	17
TOTAL	234	142

Key:

Dairy Research Institute (DRI); Tea Research Institute (TRI)
Biotechnology Research Institute (BRI); Coffee Research Institute (CRI)
Food Crops Research Institute (FCRI); Genetic Resources Research Institute (GeRRI).

After attaining the sample size of researchers of each research institute as shown on Table 4.3, a list of researchers was obtained from every research institute and used as a sample frame. Therefore, simple random sampling of researchers was selected from the sampling frame in order to attain the number of researchers required in every agricultural research institute. By so doing they were selected unbiased in the survey technique. Bryman, Teevan and Bell (2009) define simple random sampling as a subset of a statistical population in which each member of the subset has an equal probability of being chosen. The motivation behind the use of simple random sampling was to remove bias from the selection procedure and achieve representative sample (Gravetter and Forzano, 2011). All researchers sampled from each institute were reached at their place of work with prior appointment.

4.7 Data collection techniques

Data collection is the systematic approach to gathering and measuring information from a variety of sources to get a complete and accurate picture of an area of interest (McLaughlin, 2016).

Moreover, the current study used a variety of instruments in data collection which includes self-administered questionnaire, semi-structure interview and documents review. The instruments provided precise and adequate data relevant to the research problem of the study. Wegner (2000) adds that the choice of the instrument influences the quality, quantity and type of data to be gathered and selection analyses to be used. The semi-structured interview was aimed at collecting qualitative data from the directors of research institute, heads of research, heads of IT and librarians of agricultural research institutes. Self-administered questionnaire was aimed at collecting quantitative data from the researchers of agricultural research institutes. Documents review was aimed at collecting qualitative data for the study.

4.7.1 Interviews

Kvale (1996) defines interviews as an interchange of views between two or more people on a topic of mutual interest. Interviews can be used to collect both quantitative and qualitative data. While quantitative researchers use a formal, structured interview to collect precisely the data they need, qualitative researchers use a less structured and non-directive interview to explore a broader range of factors, including the thoughts and feelings of study participants (Northey, Tepperman and Albanese, 2012). Saunders, Lewis and Thornhill (2012); and Cohen, Manion and Marrison (2007) identified three categories of interviews namely structured (use questionnaires based on a predetermined and standardized or identical set of questions), semi-structured (referred to as qualitative research interview where the researcher will have a list of themes and possibly some key questions to be covered) and unstructured interviews (referred to as qualitative research interview/in-depth interview where there is no predetermined list of questions to work through but the interviewer needs to have a clear idea about the aspects). The interviewer must establish a rapport with the study participant before he/she will open up.

The study adopted semi-structures interviews to collect data from the directors of research institute, heads of research, heads of IT and librarians of agricultural research institutes. The said respondents were selected due to their experience in research data management especially RDM legal and policy frameworks, technical infrastructure, data capture, appraisal, description, preservation, access, sharing, and reuse, ICT infrastructure and RDM skill and knowledge. The data collected from semi-structured interviews were qualitative in nature (King, 2004) and used interview schedule (See appendices 1, 2 and 3 respectively).

The use of semi-structured interviews in this study was based on its strengths which are its flexibility and adaptability, ensuring a high rate of response, control of interview situations, recording of spontaneous and unintended responses and provision of in-depth data to meet specific objectives of the study (Mugenda and Mugenda, 1999). Generally, the interview questions covered themes such as: the availability or absence of legal, policy and regulations affecting RDM; research data capture, appraisal, description, preservation, accessibility and reuse; RDM knowledge, skills and training; level of ICT preparedness; and collaborative partnerships influencing RDM.

Related studies that have utilized the same method of data collection are Van Wyk and Van der Walt (2014) in a study of research data management at University of Pretoria' with fifty two (52) interviews conducted. Another study by Jao et al., (2015) on 'Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya conducted interviews on sixty (60) respondents.

4.7.2 Self-administered questionnaire

In a self-administered questionnaire, the respondents read the questions, interpret what is expected and then write down the answer. Kirklees Council (n.d) defines a self-administered questionnaire as a tool for collecting and recording information about a particular issue of interest. Questionnaires are usually associated with quantitative research which tend to be used for descriptive or explanatory research and should always have a definite purpose that is related to the research problem. Self-administered questionnaires on equal measure are qualitative in nature since it is about gaining of an understanding of opinions and motivations for human behavior. Wilson and McLean (1994) observed that questionnaire is a widely used and useful instrument for collecting survey information, providing structured, often numerical data, being able to be administered without the presence of the researcher, and often being comparatively straight forward to analyse. More importantly, the key factors in the construction of a questionnaire are the relevance of the questions to the objective of the study and to the individual respondent (Leedy, 1997).

According to Outsource2india (2016), there are two types of questions in a questionnaire design: the open format questions-open-ended questions (Don't have predetermined set of responses and the respondent is free to answer what is right) and closed format questions (where respondents are restricted to choose among any of the given multiple choice answers). Saunders, Lewis and

Thornhill (2012) categorized the questionnaires into two types namely the self-completed questionnaires and the interviewer-completed questionnaire.

The study adopted self-completed questionnaires to collect data from researchers of agricultural research institutes. The choice of a questionnaire as one of the data collection instruments for these particular respondents was informed by the fact that they are widely dispersed among the six (6) Kenya's agricultural research institutes distributed across the country. The questionnaire therefore enabled the researcher to collect data from this large number of respondents who are widely spread geographically in a cost effective manner. Kothari (2004); Pickard (2007); and Bryman, Teevan and Bell (2009) share similar sentiments on reasons for using questionnaire: data can be harvested from large sample and thus the results can be made more dependable and reliable; save cost; respondents have adequate time to give well thought out answers in the absence of researcher's effects; and anonymity can be offered as well as confidentiality.

The questionnaire used was semi-structured comprising of open-ended questions (unstructured) which constituted 15 percent of the total questions and closed-ended (structures) questions which constituted 85 percent of the total questions (see appendix 4). The data collected required very specific responses and therefore included questions that required a 'yes' or 'no' responses while others were subjected to a Likert scale or some form of quantitative (see appendix 4). Neuman (2000) asserts that the disadvantages of questionnaire can be reduced by mixing open-ended and closed-ended questions. In this study all efforts were made to effectively administer the questionnaire to the respondents and retrieve it reliably. This was done through personal contact with the respondents (researchers). The sections were organized thematically covering: the availability or absence of legal, policy and regulations affecting RDM; research data capture, appraisal, description, preservation, accessibility and reuse; RDM knowledge, skills and training; level of ICT preparedness; and collaborative partnerships influencing RDM.

Related studies that have adopted similar methods of data collection include among others: Buys and Shaw (2015) study on 'Data management practices across an institution: survey and report', and Chiliswa and Mutuku (2015) study on 'Building open data infrastructure and strategies for effective citizen engagement: case studies of Kenya and Uganda'.

4.7.3 Documents review

Documents review is a systematic procedure for reviewing or evaluating documents both printed and electronic materials (Bowen, 2009). Corbin and Strauss (2008) posit that documents review requires that data be examined and interpreted in order to elicit meaning, gain understanding and develop empirical knowledge just like other analytical methods in qualitative research. Documents review can serve a variety of functions (Bowen, 2009):

a) Documents can provide data on the context within which research participants operate;
b) Information contained in documents can suggest some questions that need to be asked and situations that need to be observed as part of the research;
c) Documents provide supplementary research data. Information and insights derived from documents can be valuable additions to a knowledge base;
d) Documents provide a means of tracking changes and developments; and
e) Documents can be analyzed as a way to verify findings or corroborate evidence from other sources.

Reviewed documents included strategic plans, legal and policy documents, annual reports, project reports on agricultural research or related area, brochures, newsletters, manuals, organizational charts among others. Perusing through the documents provided good insight and background information about the research institutes surveyed in terms of their vision, mission, objectives, core business and information on the restructuring of KARI to KALRO.

By triangulating data from documents review and interviews, the researcher attempted to provide a confluence of evidence to enhance credibility and reduce the impact of potential biases that can exist in a single study (Eisner, 1991). In sum, documents review provided background and context, supplementary data, a means of tracking change and development, and verification of findings from other data sources.

4.8 Data collection procedures

This study applied mixed methods that enabled the researcher to collect quantitative data from a large sample of researchers from Kenya’s agricultural research institutes and also qualitative data from directors of institutes, heads of research, heads of IT and librarians. The use of mixed methods ensures that data collected through one method can be validated using the other method (Creswell and Plano Clark, 2011).

Prior to the commencement of data collection, the researcher prepared data collection instruments which included questionnaires and the interview schedule. These instruments were subjected to face and content validity by selecting a sample of experts in the area of study to give a critical review on the translation of the constructs in the research instrument (Kumar, 2011; and Trochim, 2006). The instruments were equally subjected to Cronbach's alpha values of above 0.7 to measure internal consistency which involves correlating the responses to questions in the questionnaire with each other (Saunders, Lewis and Thornhill, 2012). Data collection instruments were adjusted to suit the research problem.

Upon obtaining research permits and gatekeepers authorization letters (see appendix 5, 8, 10, 12), preparation for data collection exercise began. Two weeks before the commencement of data collection exercise, the researcher visited select Kenyan agricultural research institutes for an introduction session and to book appointments with the directors of institutes, heads of research, heads of IT, librarians and researchers. The respective agricultural research institutes through their research coordinating office gave the appointment dates of when they would be available for the exercise. The research coordinating office booked appointments for interview sessions with director, heads of research, head of IT and librarian and also set up meetings with researchers to give them questionnaires. The office would further make reappointments for the interview not done and questionnaires not delivered to the researchers especially those who had assignments out of the station thus making the data collection exercise more flexible which lead to a high response rate.

Some respondents consented and they were interviewed and others filled the questionnaires which were picked up later. Some of them however declined entirely to take part in the study explaining that their schedule was too tight and did not have time to attend to the questionnaire. Some of the respondents, after many reminders, honored their word but a few of them did not respond even after repeated reminders.

During the interviews, all of the discussions were audio recorded and the researcher simultaneously also made hand written notes as a back up to the audio recording. Kalpesh (2013a); and Denscombe (2007) applauds audio recording interviews because they can be played back and listened to repeatedly to get clarity, it offers a permanent record and one that is complete in terms of the speech that occurred. Additionally audio recordings allow for further verification by other researchers. Two to three interviews were done in a day depending on the availability of the respondents to be interviewed. In some cases interview appointments had to be rescheduled on numerous occasions due to the respondents' busy nature of work. The interviews done on each day were transcribed at the end of every day to avoid any confusion with other interviews. Questionnaires were collected at the research coordinating office in every research institute and even after doing so, follow ups had to be made for those respondents who had not submitted questionnaires to the office.

4.9 Data analysis strategies

The process of data analysis involves organizing the data, conducting a preliminary read through the questionnaires and responses of the interview, coding and organizing themes, representing the data and forming an interpretation of them (Creswell, 2013). The survey questionnaire and interview schedule was used to collect data as reflected in Table 4.4 which provided a summary of the mapping of research question to data sources, respondents and the data analysis strategy.

Table 4.4: Mapping research questions to sources of data and data analysis strategy

Research question	Data sources	Respondents	Data analysis strategy
How do Kenya's agricultural research institutes capture, appraise, describe, preserve, and make accessible for reuse its research data?	Interview schedule Survey questionnaire	Directors of institutes, heads of research, heads of IT and librarians. Researchers	Qualitative data: thematic analysis Quantitative data: Statistical Package for the Social Sciences (SPSS)/ thematic analysis (open ended questions)
What knowledge, skills and training are needed to capture, appraise, describe, preserve, and make	Interview schedule Survey questionnaire	Directors of institutes, heads of research, heads of IT and librarians. Researchers	Qualitative data: thematic analysis Quantitative data: SPSS/ thematic

accessible for reuse its research data?			analysis (open ended questions)
What is the level of ICT preparedness of Kenya's agricultural research institutes to effectively capture, appraise, describe, preserve, and make accessible for reuse its research data?	Interview schedule Survey questionnaire	Directors of institutes, heads of research, heads of IT and librarians. Researchers	Qualitative data: thematic analysis Quantitative data: SPSS/ thematic analysis (open ended questions)
What policies, guidelines, and regulations are available to facilitate the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes?	Interview schedule Survey questionnaire	Directors of institutes, heads of research, heads of IT and librarians. Researchers	Qualitative data: thematic analysis Quantitative data: SPSS/ thematic analysis (open ended questions)
How do collaborative partnerships influence the capture, appraisal, description, preservation, access and reuse of research data in Kenya's agricultural research institutes?	Interview schedule Survey questionnaire	Directors of institutes, heads of research, heads of IT and librarians. Researchers	Qualitative data: thematic analysis Quantitative data: SPSS/ thematic analysis (open ended questions)

Before data analysis was carried out, data was examined for accuracy in terms of legibility, consistency and completeness of responses. Additionally interview responses were cross checked if they were complete. Equally the questionnaires were checked to eliminate those that were not properly completed. Quantitative data was analyzed using Statistical Package for the Social

Sciences (SPSS) to generate descriptive and inferential statistics while the qualitative data was analyzed thematically.

4.10 Validity and Reliability

Validity and reliability are concerned with how concrete measurement is connected to constructs and more importantly in establishing the truthfulness, credibility or believability of findings. Reliability refers to whether the same results would be received if a particular measurement technique were administered several times to the same research subjects (Bryman, Teevan and Bell, 2009). Research tools can be inferred to be reliable if they are consistent, stable, predictable, and accurate. Saunders, Lewis and Thornhill (2012) posit that reliability is concerned with the robustness of data collection instruments whether or not it will produce consistent findings at different times and under different conditions.

Research methods scholars (Creswell, 2003, 2008; Saunders, Lewis and Thornhill, 2012; Kothari, 2004; Kumar, 2011; Cohen, Manion and Marrison, 2007) identified different methods of testing reliability on data collection instruments. The methods are as follows: Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals; Parallel forms reliability is a measure of reliability obtained by administering different versions of an assessment tool to the same group of individuals; Inter-rater reliability is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions; and finally internal consistency reliability is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results (two subtypes of Internal consistency reliability: Average inter-item correlation and Split-half reliability).

Validity on the other hand is defined by Smith (1991) as the degree to which the researcher has measured what he has set out to measure. Kerlinger (1973) epitomized the definition of validity by asking: “Are we measuring what we think we are measuring?” Comparably, Saunders, Lewis and Thornhill (2012) refer to validity as the extent to which data collection method or methods accurately measure what they were intended to measure. In relation to this, Phelan and Wren (n.d); Bloomberg and Volpe (2008) and Kumar (2011) classify validity into the following types: face and content validity where each question on the research instrument must have a logical link with an objective and must also cover the full range of the issue or attitude being measured; predictive validity is used to predict future or current performance-it correlates test results with another criterion of interest; concurrent validity is judged by how well an instrument compares with a second assessment concurrently done; and finally the construct validity refers to the extent to which the measurement questions actually measure the presence of those constructs you intended them to measure.

While reliability is necessary, it alone is not sufficient; conversely for a test to be reliable, it also needs to be valid. The validity and reliability of the data collected by the researcher and response rate achieved depend on the design of the questions and the structure of the questionnaire. Moreover, Saunders, Lewis and Thornhill (2012) point out that a valid questionnaire will enable accurate data that actually measure the concepts the researchers are interested in collecting, whilst one that is reliable will mean that the data are collected consistently. The use of reliability and validity are common in quantitative research rooted in positivist perspective, however they should be redefined for their use in a naturalistic approach (interpretive paradigm). Gorman and Clayton (2005) attest to this arguing that the quality of research instruments in qualitative studies cannot be judged using the statistical measure of validity and reliability.

To ensure reliability in qualitative research, examination of trustworthiness is crucial (Golafshani, 2003). In a rejoinder Lincoln and Guba (1985) pointed out that to establish the 'trustworthiness' of a study, terms such as credibility, authenticity, transferability, dependability and confirm ability should be used as a measure of reliability in qualitative research. Many researchers (Davies and Dodd, 2002; Lincoln and Guba, 1985; Mishler, 2000; Stenbacka, 2001) have developed concepts of measuring validity in qualitative research and have considered them to be more appropriate terms such as quality, rigor and trustworthiness. In this view, Patton (2002:247) advocates the use of triangulation by stating:

> …triangulation strengthens a study by combining methods. This can mean using several kinds of methods or data, including using both quantitative and qualitative approaches.

In the present study, validity and reliability were assured through the following methods:

- Validity of the research instruments was achieved through face and content validity by linking the questions and the objectives of the study (Kumar, 2011). Face validity according to Trochim (2006), looks at the operationalization of the research instrument and assess whether 'on its face' it seems like a good translation of the constructs. Trochim asserts that quality of face validity assessment can be improved considerably by making it more systematic and carefully selecting a sample of experts in the area of study to give a critical review on the translation of the constructs in the research instrument. In this case selected lecturers in the school of Information Sciences Moi University, Kenya who were acquainted with RDM, library, IT and records management were invited to critique the data collection tools of which their suggestions improved the data collection tools.

- In order to attain validity and reliability in the present study, a pilot study was carried out. A pilot study refers to feasibility studies which are small scale versions or trial runs done in preparation for the major study (Polit and Beck, 2001). The reasons for conducting a pilot study in the present study (Van Teijlingen and Hundley, 2001) was to: develop and test adequacy of research instruments, design a research protocol, and establish whether the sampling frame and technique are effective. Pilot study allowed the researcher to gauge the meaning attributed to survey questions both in interview and questionnaire form. Pre-testing of the data collection instruments was carried out at KEFRI-Marigat centre to prevent contamination which would occur if the same participants in the main study are included in the pilot study (Van Teijlingen and Hundley, 2001). KEFRI-Marigat centre was selected for a pilot study because it had similar characteristics related to agriculture. Additionally the fact that it was a research institute made it even more suitable (Check and Schutt, 2012). Interview schedules were administered to one research institute director, four heads of research, two librarians, three IT experts and questionnaires were administered to twenty two (22) researchers making a total of thirty two (32) respondents. The number was informed by Johanson and Brooks (2010) who suggests that 30 representative participants is a reasonable minimum recommended for a pilot study. Once the pilot study was complete, the outcome of the interviews and the questionnaire were assessed consulting scholars with experienced in RDM or related fields. Further, the statistician used the outcome of the questionnaires to test the reliability of the instruments using Cronbach values of above 0.7. The responses obtained were used to adjust the data collection instruments to suit the study.

- Cronbach's alpha was used in the study as a measure of internal consistency which involves correlating the responses to questions in the questionnaire with each other (Saunders, Lewis and Thornhill, 2012). Cronbach's alpha consists of an alpha coefficient with a value between 0 and 1. According to Streiner (2003) when alpha of Cronbach is .90 > α ≥ .80 it is a good coefficient, when alpha of Cronbach is .80 > α ≥ .70 it is to be acceptable , when alpha of Cronbach is .70 > α ≥ .60 it is questionable, when alpha of Cronbach is .60 > α ≥ .50 internal consistency of tools is poor.

In the present study, reliability analysis was conducted for the five scales measuring: Availability or absence of legal, policy and regulations of RDM; Research data capture, appraisal, description, preservation, access and reuse; Knowledge, skills and training needed for RDM; Level of ICT preparedness for RDM; and Collaborative partnerships influencing RDM. Results presented in Table 4.5 shows the Cronbach's alpha values for the item total correction coefficients and that all the scales were reliable in measuring the stated variables.

Table 4.5: Cronbach's alpha values

Research Question	**Number of Items**	**Cronbach's Alpha**
Availability or absence of legal, policy and regulation of RDM	20	0.907
Data capture, appraisal, description, preservation, access and reuse	146	0.709
Knowledge, skill and training needed for RDM	30	0.698
Level of ICT preparedness for RDM	7	0.703
Collaborative partnerships influencing RDM	14	0.762

- Triangulation method was equally used to achieve reliability and validity in the present study. Bogdan and Biklen (2006) assert that triangulation is a powerful technique that facilitates validation of data through cross verification from two or more sources. Data collection tools were triangulated using questionnaires, interviews and document review (See Appendices 1, 2, 3, and 4 respectively) by asking respondents similar questions on

RDM in order to address the research problem. Triangulation technique was utilized to ensure that the study was rich, robust, comprehensive and well-developed in terms of data collected (Patton, 2002).

Cox and Pinfied (2014) tested the reliability of questionnaire using Cronbach's Alpha Value in a study of 'Research data management and libraries: current activities and future priorities'. The study adopted relevant questions from existing tools with a Cronbach's alpha value of above 0.7.

4.11 Ethical considerations

Ethics refers to doing what is morally and legally right in the conducting of research. Resnik (2015) defines ethics as a method, procedure or perspective for deciding how to act and for analyzing complex problems and issues. Saunders, Lewis and Thornhill (2012) refer to ethics as the standards of behavior that guide the researcher's conduct in relation to the right of those who become the subject of the study. Resnik (2015) postulates reasons why it is important to adhere to ethical norms in research:

(a) Norms promote the aims of research, such as knowledge, truth and avoidance of error;
(b) Ethics standards promote the values that are essential to collaborative work such as trust, accountability, mutual respect and fairness since research involves a great deal of cooperation and coordination among many different people;
(c) Ethical norms in research help to ensure that researchers can be held accountable to the public;
(d) Ethical norms in research help to build public support for research; and
(e) Many of the norms of research promote a variety of other important moral and social values like social responsibility, compliance with the law among others.

Cohen, Manion and Marrison (2007) identify ethical issues that constitute a set of ethical considerations that researchers across the board should address in (/when? planning research:

(a) Informed consent;
(b) Gaining access to and acceptance in the research setting;
(c) Source of tension in the ethical debate, including non-maleficence, beneficence and human dignity, absolutist and relativist ethics;
(d) Problems and dilemmas confronting the researcher, including matters of privacy, anonymity, confidentiality, betrayal and deception;
(e) Ethical problems endemic in particular research methods;

(f) Ethics and evaluative research;

(g) Regulatory ethical frameworks, guidelines and codes of practice for research;

(h) Personal codes of practice;

(i) Sponsored research; and

(j) Responsibilities to the research community.

Whereas the above mentioned ethical issues concern procedural ethics, the researcher had to consider the research purpose, contents, methods, reporting and outcomes complied with ethical principles and practices. The study complied with UKZN research ethical guidelines (see appendix 5). In addition, a research permit was sought from NACOSTI in Kenya (see appendices 7 and 8 respectively). Further permission was sought from the Kenya agricultural research institutes where the study was undertaken (see appendices 9, 10, 11, and 12 respectively). In addition, consent was sought from the respondents accepting to participate in the study voluntarily (see Appendix 6). The respondents were assured of privacy, confidentially and anonymity. The respondents were advised that they could withdraw from the study at any stage if they so wished without any sanctions.

Summary

The chapter discusses the research methodology used in the study. It charted out the overall research about the study and more importantly gave a road map by presenting the research design that addressed the research problem of the study. The themes covered in this chapter consist of: research paradigm, research methods, research designs, population of the study, sampling procedures, data collection techniques, data collection procedures, data analysis strategies, validity and reliability and ethical consideration. Before data analysis was carried out, data was examined for accuracy in terms of legibility, consistency and completeness of responses. Additionally interview responses were cross checked if they were complete. Similarly the questionnaires were checked to eliminate those that were not properly completed. Quantitative data was analyzed using SPSS to generate descriptive and inferential statistics while the qualitative data was analyzed thematically.

The next chapter on Data Analysis and Presentation of the findings presents the results of the empirical study.

CHAPTER FIVE

DATA ANALYSIS AND PRESENTATION OF FINDINGS

5.1 Introduction

Data analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data (Kalpesh, 2013b). The essential component of data analysis involves explanations, comparisons, predictions and exploration of inter-relationships between variables (Loughborough University, nd). In essence, Data analysis brings order, structure and meaning to the mass of collected data in order to showcase the empirical findings in an attempt to answer the research questions addressed by the study. Furthermore, data analysis provides an explanation of various concepts, theories, frameworks and methods used (Kalpesh, 2013b; Johnson, 2011).

This study applied mixed methods that enabled the researcher to collect quantitative data from a large sample of researchers from Kenya's agricultural research institutes and also qualitative data from directors of institutes, heads of research, heads of IT and librarians. The use of mixed methods ensures that data collected through one method can be validated using the other method (Creswell and Plano Clark, 2011). The purpose of this study is to examine Research Data Management (RDM) in Kenya's agricultural research institutes with the view to propose interventions to improve management, sharing and reuse of agricultural research output.

The respondents were composed of directors of institutes, heads of research, heads of IT, librarians, and researchers from six (6) of Kenya's agricultural research institutes namely: research institute A, research institute B, research institute C, research institute D, research institute E, and research institute F. Anonymous names are given to sampled agricultural research institutes to maintain confidentiality. Further, the study was underpinned by Data Curation Centre (DCC) lifecycle model (Higgins, 2008) and Community Capability Model (CCM) framework (Lyon et al., 2012).

The quantitative data collected through questionnaires were prepared for analysis through coding before being entered into SPSS program. Thereafter frequency tables, graphs and statistics were generated and used to present, describe and examine the results. This chapter is structured into several sections. In the first section, a description of the demographic profiles of the respondents is presented. This is followed by a descriptive analysis of the study variables in conjunction with the results of thematic analyses from survey questionnaires and in-depth interviews conducted with the directors of research institutes, heads of research, heads of IT and librarians. The results are presented using the research questions as the organizing framework.

5.2 Response rate

A perfect representative sample is one that exactly represents the population from which it is taken. Essentially, response rate refers to the number of people who completed interviews and questionnaires issued, divided by the number of people in the sample, usually expressed in the form of a percentage. Babbie and Mouton (2001) assert that a response rate of 50 percent is adequate for analysis and reporting, while a response rate of 60 percent is good and a response rate of 70 percent is very good. This was in agreement with the assertions of McLaughlin, Bush and Zeeman (2016) and Bryman (2012) stating that the acceptable response rate should be at least 60%. They further pointed out that response rates can be ranked, 60-69% acceptable, 70-85% very good and 85% and over excellent.

Related studies have attained higher response rates. For example, a study by Pinfield, Cox and Smith (2014) on RDM and libraries: relations, activities, drivers and influences involved 26 interviews of library practitioners from different institutions in the UK and attained a response rate of 77%. The interviews were a follow-up of the same project which began with an online survey of 156 respondents of UK academic libraries of which attained a response rate of 71%. In the current study, the questionnaires were administered to 142 researchers. The questionnaires returned were used to analyze data of 124 representing 87% response rate. Similarly for the interviews, the response rate was 80% as indicated on Table 5.1. These were high and acceptable response rates as recommended by Babbie and Mouton (2001), McLaughlin, Bush and Zeeman (2016) and Bryman (2012). Groves and Peytcheva (2008), assert that high response rates are preferable to reduce the risk of non-response bias and to ensure that the sample is representative.

Table 5.1: Interviews and questionnaire response rates (N=41 and 142; n=33 and 124)

Target Group	Sample size (N)	Response rate (n)	Response rate in percentages (%)
Interviews			
Directors of Institutes	6	6	100.0

Heads of Research	23	18	78.0
Heads of IT	6	4	67.0
Librarians	6	5	83.0
Interviews response rate	**41**	**33**	**80.5**
Questionnaires			
Research institute A	40	35	87.5
Research institute B	19	16	84.2
Research institute C	24	21	87.5
Research institute D	16	14	87.5
Research institute E	17	15	88.2
Research institute F	26	23	88.5
Researchers response rate	**142**	**124**	**87.0**

Source: Field data (2017).

Interviews: N=41; n=33. **Questionnaire**: N=142; n=124.

The high response rate was attributed to the persistence and follow ups by the researcher.

5.3 Background information of the respondents

The respondents who participated in this study were asked to state the name of the institution where they were working. The number of researchers who responded to the survey amounted to 124 and the number that responded to interviews was 33. The interviews were administered to the directors of institutes, heads of research, heads of IT and librarians. The results are presented in the Table 5.2.

Table 5.2: Biographical information of the researchers (Questionnaires) (n=124)

Bio-graphical information	Categories	Frequency	Percentage
Name of Research institute	ARI A	35	28.2
	ARI B	16	12.9
	ARI C	21	16.9
	ARI D	14	11.3
	ARI E	15	12.1
	ARI F	23	18.5
	Total	**124**	**100.0**
Gender			
	Male	68	54.8
	Female	56	45.2
	Total	**124**	**100.0**

Age			
	26-30	14	11.3
	31-35	19	15.3
	36-40	30	24.2
	41-45	19	15.3
	46-49	19	15.3
	50-above	23	18.5
	Total	**124**	**100.0**
Field of Specialization			
	Records Management	10	8.1
	Agricultural Pathology	17	13.7
	Agricultural economics	14	11.3
	Agricultural data analysis	8	6.5
	Agricultural Economics	6	4.8
	Breeding Soil science	20	16.1
	Laboratory technology	10	8.1
	Dairy science	2	1.6
	Biochemistry	24	19.4
	Seed science and technology	4	3.2
	Zoology/entomology	4	3.2
	Molecular Biotechnology	5	4.0
	Total	**124**	**100.0**
Experience	less than 1 year	8	6.5
	1-5 years	27	21.8
	5-10 years	39	31.5
	10-15 years	20	16.1
	above 15 years	30	24.2
	Total	**124**	**100.0**
Highest educational level			
	Diploma	7	5.6
	Higher diploma	11	8.9
	Bachelor’s degree	44	35.5
	Master’s degree	41	33.1
	PhD	21	16.9
	Total	**124**	**100.0**

Source: Field data, 2017; **ARI**- Agricultural research institute

The findings indicates that 35(28.2%) of the respondents were working in research institute A, while 23(18.5%) were working in research institute F. There were 21(16.9%) who were working in Research institute C, 16(12.9%) were working in research institute B whereas 14(11.3%) were working in research institute C. Only 15(12.1%) were working in research institute E. It was also revealed that 68(54.8%) of the researchers were male while 56(45.2%) were female. These results suggest that almost an equal number of the researchers from either sex participated in this study. Besides, 30(24.2%) of the researchers were aged 36-40 years, while 23(18.5%) were above 50 years old. Another 19(15.3%) were aged 31-35 years, 41-45 years, and 46-49 years respectively. There were 14(11.3 %) who were aged 26-30 years old. This result implies that the majority of the respondents were below 50 years but above 26 years of age.

Furthermore, 10(8.1%) of the participants had their field of specialization in records management, 17(13.7%) specialized in agricultural pathology whereas 14(11.3%) had their specialization in agricultural economics. The majority of the respondents specialized in Biochemistry 24(19.4%), followed up by Breeders at 20(16.1%). However, another 4(3.2%) specialized in seed science and technology. In addition the results show that 8(6.5%) and 6 (4.8%) of the respondents specialized in environmental data analysis and land economics respectively. The lowest number 2(1.6%) of the respondents specialized in animal production because only one institute researching on livestock was sampled in the study. Moreover, 10(8.1%) of respondents specialized in laboratory technology while 5(4.0%) specialized in molecular biotechnology.

The findings also indicated that 39(31.5%) were in their current institute of affiliation for 5-10 years while 30(24.2%) had worked for more than 15 years. The results show that 27(21.8%) of the respondents had worked for 1-5 years while 20(16.1%) had been in the same station for 10-15 years. There were 8(6.5%) of the researchers who have been in the station for less than 1 year. The results also reveal that 44(35.5%) of the respondents were bachelor's degree holders and 41(33.1%) were masters' degree holders. Another 21(16.9%) were PhD holders whereas 11(8.9%) were higher diploma holders and only 7(5.6%) were diploma holders.

The rest of the chapter is organized based on themes of the research questions (see section 5.1) as follows: The availability or absence of legal, policy and regulations affecting the capture, appraisal, description, preservation, access and reuse of research data; Research data capture, appraisal, description, preservation, accessibility and reuse; RDM knowledge, skill and training requirements needed to capture, appraise, describe, preserve, access and reuse; Level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse; and Collaborative partnerships influencing the capture, appraisal, description, preservation, access and reuse. Similarly the findings from interviews and questionnaires are collated and also presented under the same themes.

5.4 Legal, policy and regulatory framework for RDM in Kenya's agricultural research institutes

One of the underpinning models for the study in relation to legal, policy and regulations governing RDM is the Community Capability Model (CCM) (Lyon et al., 2012:3) described in chapter two (2) of this thesis. Similarly, DCC lifecycle model (Higgins, 2008) which also informs the study is equally relevant in this research question because it can be used to ensure that processes and policies are adequately documented for all the required stages of data curation after being identified and planned. Legal and policy frameworks provides clarity on what is expected from the agricultural research institute and who is responsible for which activities accordingly having a critical influence on data capture, appraisal, description, preservation, access and reuse. Consequently, RDM legal, policy and regulations framework are needed to guide the practices of data sharing, intellectual property, human resource capability, technical infrastructure, ethical issues and open access (Anderson, 2004). The Cronbach's Alpha values for this question were 0.907 for items in the researcher's questionnaire. This suggested a high internal validity of the test items. Questions B (1-2) of the questionnaire for researchers (Appendix 4), questions B (1-6) of the interview schedule for directors of institute (Appendix 1), questions B (1-5) of the interview schedule for heads of research (Appendix 2), and questions B (1-4) of the interview schedule for heads of IT and librarians (Appendix 3) addressed this research question. The results are presented in the following sections 5.4.1 and 5.4.2.

5.4.1 Legal framework governing RDM

The study sought to determine whether there were legal frameworks that governed RDM in the institutes. The responses are shown in fig. 5.1.

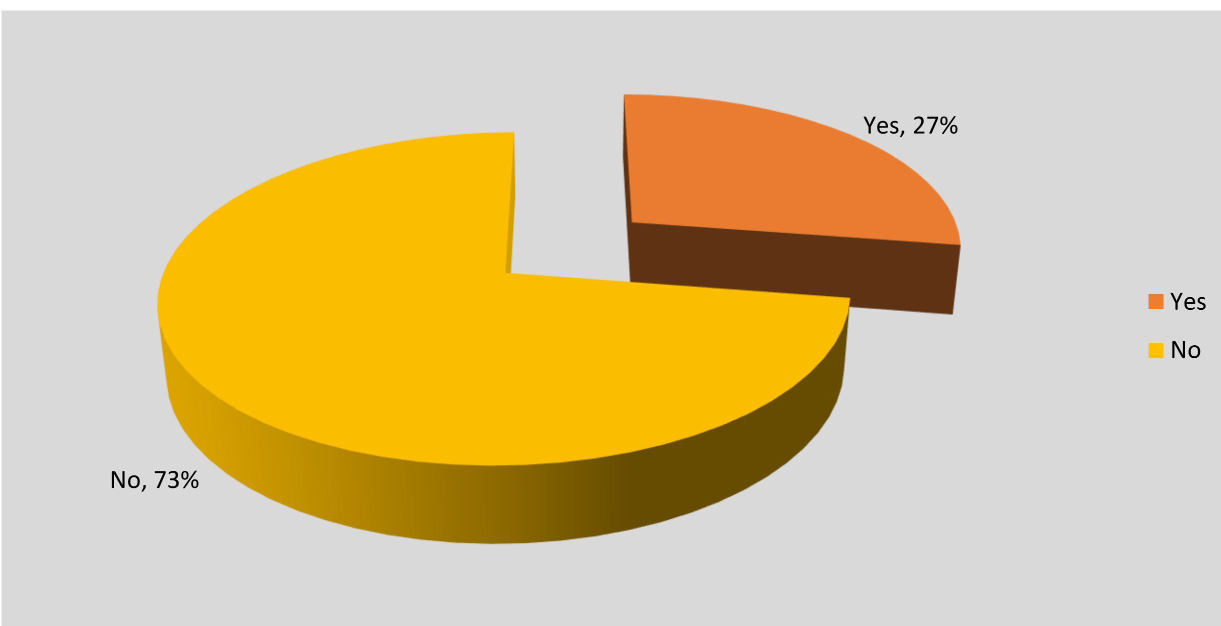

Figure 5.1 Legal framework governing RDM (n=124)

The findings show that 90(72.6%) of the researchers stated that there was no legal frameworks that govern RDM in the institutes. Majority of respondents interviewed indicated that the legal frameworks that govern RDM do not exist because this requirement is not incorporated in the KALRO Act (No.17 of 2013). KALRO Act (No.17 of 2013) states:

> *"An Act of Parliament to provide for the establishment and functions of the Kenya Agricultural and Livestock Research Organization; to provide for organs of the Organization; to provide for the co-ordination of agricultural research activities in Kenya, and for connected purposes".*

It can be noted from the Act that issues of RDM are not addressed.

Furthermore, they indicated that they were not aware of any stand-alone legal frameworks that govern RDM in all research institutions in Kenya. On the same note, the respondents acknowledged the presence of Intellectual Property (IP) policy and International Standards Organization (ISO) - ISO 9001: 2008 certification which has gone a long way in assisting them in RDM functions and activities. The remaining 34(27.4%) agreed that there was a legal framework governing RDM which was enshrined in ISO 9001:2008 certification. The respondents were not able to differentiate between the ISO certification and the legal frameworks. The ISO is an external international standard, while legal framework is a national instrument. According to responses of two heads of research HR2 and HR5 from Research institute B and F respectively concerning legal framework, HR2 they stated that:

> *There is no formal legal framework, what is there is just the public servants guidelines which talks about ethical ways of conducting research by public servants and also ways of keeping data. There are also procedures for conducting research which provide guidelines including intellectual property policy.*

According to the testimony of HR5:

> *The institute follows ISO guidelines as guiding principles for data management. Before ISO came into operation everyone was conducting research in t,heir own way using notebooks. Also the Standard Operating Procedures (SOPs) that guide researchers in undertaking research in most research institutes has played a role in educating researchers on the ethical code conduct when doing research projects.*

ISO has equally played a crucial role in RDM in delaying down procedures of research data capture, labeling, storing and others.

5.4.2 Policies and regulations available to facilitate RDM

The respondents were asked to state the policies and regulations that were available to facilitate the capture, appraise, description, preservation, access and reuse of research data. There were 20 items measuring this variable. The responses were coded as 1 = strongly agree, 2 = agree, 3 = undecided 4=disagree, and 5 = strongly disagree. While scoring the questionnaires the highest possible score for each item on the Likert scale was 1.0 points and the lowest was 5.0. The highest possible mean score for a respondent was 1.0 and the lowest was 5.0. The midpoint was taken to be 3.0 and this was used to categorize responses as either "agree" or "disagree". For each item a mean and standard deviation were calculated. The responses are presented on Table 5.3.

Table 5.3: Policy and regulations available to facilitate RDM (n=124)

Statement	Mean	Standard deviation
RDM policy	2.0806	.69372
The research institute has RDM policy governing:		
Data capture	2.0000	.79633
Data appraisal	2.3065	.97261
Data description	2.0726	.78785
Data preservation	2.0403	.84005
Data access	2.0645	.83366
Data use and reuse	2.0887	.79646
Data sharing	2.1210	.87970
Knowledge/skill/training	2.4274	1.01351
Technical infrastructure	2.1290	.91924
Collaborative Partnerships	2.0403	.89624
The institute owns the right to research data	1.7823	.76041
The Researcher owns the right to data created	1.7016	.70974
Intellectual property affects RDM	1.9839	.90152
There is quality assurance and control measures in place during:		

Research data capture	1.8306	.69518
Research data appraisal	2.1290	.90138
Research data description	1.9355	.82385
Research data preservation	1.9597	.85918
Research data access	1.9194	.73911
Research data reuse	1.9677	.76441

Source: Field data (2017)

The findings show that the majority of respondents agreed that the research institute generally had RDM policy (Mean=2.0806, SD=0.69372). Specifically, the respondents asserted that the research institute had RDM policy on research data capture (Mean=2.0000, SD=0.79633), research data appraisal (Mean=2.3065, SD=0.97261), research data description (Mean=2.0726, SD=0.78785), research data preservation (Mean=2.0403, SD=0.84005) and RDM policy governing research data access (Mean=2.0645, SD=0.83366).

The respondents generally agreed that there was RDM policy governing research data use and reuse (Mean=2.0887, SD=0.79646), research data sharing (Mean=2.1210, SD=0.87970), knowledge, skills and training (Mean=2.4274, SD=0.01351), technical infrastructure (Mean=2.1290, SD=0.91924) and collaborative partnerships (Mean=2.0403, SD=0.89624). The findings also show that most research institutes where the study was done own the rights to research data (Mean=1.7823, SD=0.76041). Equally, the researchers own the right to research data created (Mean=1.7016, SD=0.70974). It is also shown that the majority of respondents stated that intellectual property affects RDM (Mean=1.9839, SD=0.90152).

The respondents were also asked to state whether there were quality assurance and control measures in place during data capture, appraisal, description, preservation, access and reuse. The results indicated in Table 5.3 show that the respondents agreed that there were quality assurance and control measures in place during research data capture(Mean=1.8306, SD=0.69518), research data appraisal(Mean=2.1290, SD=0.90138), research data description(Mean=1.9355, SD=0.82385) and research data preservation(Mean=1.9597, SD=0.85918). Further, the result shows that there was quality assurance and control measures in place during research data access (Mean=1.9194, SD=0.73911) and research data reuse (Mean=1.9677, SD=0.76441).

Results from the interviews reveals varied responses concerning policy and regulations available to facilitate RDM in the research institutes where the study was done. There were institutes where there were neither policies nor regulations, whereas in majority of the institutes, there was evidence of policy and regulation governing RDM. Heads of ICT and librarians who were interviewed said that there were policies and regulations to facilitate access, reuse and sharing of research data. However, there were institutes that were still working on the policies and regulations aligned to RDM.

Concerning challenges in data capture, appraisal, description, preservation, access and reuse, heads of IT and librarians who were interviewed stated that they were excluded in administration of the policies and regulations concerning RDM. Heads of IT and librarians lamented that research data access, reuse and sharing has greatly been hampered because of their exclusion in RDM, furthermore policies on appraisal, description and preservation are inadequate. In addition, plagiarism of research data collected was also stated as the main challenge facing researchers in data access and sharing. Librarian [L1] from research institute A interviewed stated:

> *"The role the library plays in RDM is not recognized and it is not included in editorial committee and setting up of policies and regulations governing research data. The Library is not aware of what other departments in their area of research are doing and the library should be present to capture the proceeding………. Plagiarism is also a problem since most researchers are scared of putting their research data in the library because other researchers will copy other people's work. Scientific papers published are also not reaching the library since scientists are weary of handing over their work to the library. If there was RDM policy in KALRO guiding RDM stakeholders in the institutes from capturing data to access, sharing and reuse, which could be a big step in managing data. Library and IT department could play major role"*

These results suggest that there are no clear policies and regulations for RDM. In addition there mistrust seems to exist among departments and researchers. With regard to Intellectual Property (IP), quality assurance and control measures, the directors and heads of research interviewed acknowledged that Kenya's agricultural research institutes had a well stipulated IP policy and were compliant with ISO 9001: 2008 certification. ISO 9001: 2008 specifies requirements for a quality management system which should be a strategic decision of an organization. In essence, quality management systems should be influenced by: its organizational environment, its varying needs, its particular objectives, the products it provides, the processes it employs, and its size and organizational structures.

From ISO 9001: 2008, KALRO draws several procedure including: procedure for provision of services, Standard Operating Procedures (SOPs), packaging and dissemination of technology, information security and backup to mention a few. All agricultural research institutes country wide must conform to ISO 9002: 2008 standards. However, directors of institutes and heads of research noted that there was a lack of risk/disaster management policy for RDM. According to the testimony of one of the directors [D2] from Research Institution C on risk/disaster management policy for RDM:

> *There is no policy on risk or disaster management as such but the institute trains staff annually on risk and disaster management well as information on password leakages and information landing in wrong hands. It's normally on* Quality Management *System (QMS) that all stakeholders in agricultural research institute should be oriented on disaster management. Actually the institutes should relook and consider having disaster management policy because research data needs to be taken care of because it is an asset to the institute.*

5.5 Research Data Management in Kenya's agricultural research institutes

The Data Curation Centre (DCC) lifecycle model (Higgins, 2008: 136) in chapter two (2) promotes a lifecycle approach to the management of digital material to enable their successful data curation. According to the model, data curation includes data capture, appraisal, description, preservation, access, reuse and transformation of research data. The model underpins the theme above (5.5), by advocating for maintenance of authenticity, reliability, integrity and usability of digital materials which in return ensures quality of RDM. Equally, Community Capability Model (CCM) framework (Lyon et al, 2012) informs the theme (5.5) by advocating in growing the capability of RDM communities to perform data-intensive research through comprehensive and coherent legal and policy frameworks, qualified human resource, up to date technical infrastructure and enhancing collaborative partnerships to facilitate access and sharing. The Cronbach's Alpha values for this question were 0.709 for items in the researcher's questionnaire. This suggested a high internal validity of the test items. In this respect, the study sought to establish how research data is captured, appraised, described, preserved, accessed and reused in Kenya's agricultural research institutes. Questions C (1-2) of the questionnaire for researchers (Appendix 4), questions C (1-5) of the interview schedule for directors of institute (Appendix 1), questions C (1-8) of the interview schedule for heads of research (Appendix 2), and questions C 1-6) of the interview schedule for heads of IT and librarians (Appendix 3) addressed this research question. The findings are presented in the following sub-titles:

5.5.1 Capturing research data

In order to understand how research data was captured, researcher's responses are summarized in Table 5.4

Table 5.4: Methods of capturing research data (n=124)

Respondent	Responses
R4	*Using questionnaires and interviews to collect data and analyse using various statistical packages*
R13	*Capture data using cameras, Audio, observations, recordings, field notebook*
R27	*Data is captured by collecting tissue samples, doing laboratory tests, classifying and coming up with data.*
R56	*From field experiments data is collected, also seed samples are collected in the field, they are classified using pre-determined data sheet, taken to laboratory for testing of which results are analysed and data released.*
R79	*Data is also captured through laboratory experiments of which they are analysed, use of Geographic Information Systems (GIS).*
R98	*Through the use of seminars, conferences, workshops and journal publications, new data is generated.*

N/B	*Data collected must meet the minimum set criteria by the agricultural research institute before being analysed. These criteria are documented and adherence to the same is overseen by the Heads of research.*

Source: Field data (2017)

The responses given above in (Table 5.4) summarize how research data is captured. It was evident that the commonly used method of capturing data was via the use of a questionnaire, interviews, audio recordings, cameras, GIS, laboratory experiment and field experimentation. Subsequently, the captured research data was analyzed and disseminated to the researcher or heads of research for consultation or use.

5.5.1.1 Formats for capturing research data

The respondents were asked to state the formats they captured or generated their research data with. The findings are presented in Table 5.5.

Table 5.5: Formats for capturing research data (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Audio	9	7.3	51	41.1	26	21.0	25	20.2	13	10.5	124	100.0
Images	30	24.2	71	57.3	14	11.3	6	4.8	3	2.4	124	100.0
Spreadsheet	39	31.5	59	47.6	18	14.5	1	0.8	7	5.6	124	100.0
Video	26	21.0	58	46.8	21	16.9	9	7.3	10	8.1	124	100.0
Data-statistical	44	35.5	61	49.2	15	12.1	1	0.8	3	2.4	124	100.0
Database	32	25.8	74	59.7	10	8.1	6	4.8	2	1.6	124	100.0
Scanned document	34	27.4	59	47.6	19	15.3	7	5.6	5	4.0	124	100.0
Web	30	24.2	49	39.5	27	21.8	12	9.7	6	4.8	124	100.0
Computer Aided Design (CAD)	16	12.9	51	41.1	28	22.6	10	8.1	19	15.3	124	100.0
Geographical Information System (GIS)	22	17.7	43	34.7	35	28.2	12	9.7	12	9.7	124	100.0
Data XML	12	9.7	52	41.9	39	31.5	4	3.2	17	13.7	124	100.0

Source: Field data (2017)

The results suggest that the majority of respondents captured or generated their research data in images 101(81.5%), spreadsheets 98(79.1%), video 84(67.8%), data statistical (SAS, SPSS) 105(84.7%) and database 106(85.5%). There were 93(75%) of the respondents who stated that they generated their data in scanned document format, whereas 79(63.7%) of the respondents stated that they generated research data in web format. However, of the same group of respondents 12(9.6%) and 18(14.5%) disagreed that they do not capture data on scanned documents and web respectively. Another 67(54%), 65(52.4%), 64(51.6%) and 60(48.4%), of the respondents respectively generated their research data in the form of computer aided design, GISs, data XML and audio format respectively. On the contrary, 35(28.2%) of the respondents were undecided on capturing data using GIS. However, another 24(19.4%) disagreed that they do not use GIS to capture data.

The same sentiments were echoed by the directors of the institutes and heads of research in emphasizing that audio and video recordings were used in interviewing farmers, analysis of data was done using SPSS and Gen STAT and computer programs that assist in capturing data. One of the directors (D2) from Research institute B states:

> *In Sustainable ecosystems and environmental management, meteorological data is captured using thermometers, rain gauge and other tools both in hard and soft copy. In this regard, captured data generates weather reports which assist the institutes in planning its research projects and also the surrounding communities benefit from the same data. In socio-economics, during open and field day, there is the use of questionnaires (hard copies), interviews, videos and audio recording to capture data from farmers, consumers and other stakeholders on the product(s) of the research institute. The primary data captured is of great value to the institute because it will gauge the quality of the products according to ISO certification. At times GIS is used to capture, store and analyze geographical data especially in research institutes that have research projects that depend much on geographical data. In addition, use of computer like email, internet and others plays a role in capturing data.*

5.5.1.2 Approximate amount of research data generated

The respondents were also asked to approximate the amount of research data their institute had generated in every research project. The responses are shown in Table 5.6.

Table 5.6 Cross tabulation of agricultural research institutes and approximate amount of research data generated in every research project (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
1-500GB	**A**	13	10.5	10	8.1	10	8.1	1	0.8	1	0.8	**35**	**28.2**
	B	7	5.6	1	0.8	7	5.6	1	0.8	0	0.0	**16**	**12.9**
	C	3	2.4	6	4.8	6	4.8	3	2.4	3	2.4	**21**	**16.9**
	D	3	2.4	5	4.0	3	2.4	0	0.0	3	2.4	**14**	**11.3**
	E	5	4.0	7	5.6	2	1.6	0	0.0	1	0.8	**15**	**12.1**
	F	0	0.0	13	10.5	4	3.2	0	0.0	6	4.8	**23**	**18.5**
500-1000GB	**A**	10	8.1	11	8.9	13	10.5	0	0.0	1	0.8	**35**	**28.2**
	B	4	3.2	3	2.4	9	7.3	0	0.0	0	0.0	**16**	**12.9**
	C	4	3.2	8	6.5	7	5.6	0	0.0	2	1.6	**21**	**16.9**
	D	3	2.4	5	4.0	3	2.4	1	0.8	2	1.6	**14**	**11.3**
	E	7	5.6	6	4.8	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	0	0.0	13	10.5	0	0.0	4	3.2	6	6.8	**23**	**18.5**
1-500TB	**A**	11	8.9	9	7.3	13	10.5	2	1.6	0	0.0	**35**	**28.2**
	B	4	3.2	2	1.6	9	7.3	1	0.8	0	0.0	**16**	**12.9**
	C	6	4.8	6	4.8	7	5.6	0	0.0	2	1.6	**21**	**16.9**
	D	3	2.4	4	3.2	3	2.4	1	0.8	3	2.4	**14**	**11.3**
	E	7	5.6	6	4.8	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	2	1.6	15	12.1	3	2.4	0	0.0	3	2.4	**23**	**18.5**
500 – 1000TB	**A**	10	8.1	10	8.1	13	10.5	1	0.8	1	0.8	**35**	**28.2**
	B	3	2.4	3	2.4	9	7.3	0	0.0	1	0.8	**16**	**12.9**
	C	5	4.0	6	4.8	8	6.5	0	0.0	2	1.6	**21**	**16.9**
	D	3	2.4	4	3.2	3	2.4	1	0.8	3	2.4	**14**	**11.3**
	E	7	5.6	6	4.8	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	2	1.6	15	12.1	3	2.4	0	0.0	3	2.4	**23**	**18.5**

1-500 PB	**A**	11	8.9	9	7.3	13	10.5	1	0.8	1	0.8	**35**	**28.2**
	B	4	3.2	2	1.6	9	7.3	0	0.0	1	0.8	**16**	**12.9**
	C	4	3.2	7	5.6	8	6.5	0	0.0	2	1.6	**21**	**16.9**
	D	3	2.4	4	3.2	3	2.4	1	0.8	3	2.4	**14**	**11.3**
	E	7	5.6	6	4.8	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	2	1.6	15	12.1	3	2.4	0	0.0	3	2.4	**23**	**18.5**
>500PB	**A**	13	10.5	9	7.3	13	10.5	0	0.0	0	0.0	**35**	**28.2**
	B	5	4.0	2	1.6	9	7.3	0	0.0	0	0.0	**16**	**12.9**
	C	4	3.2	7	5.6	8	6.5	0	0.0	2	1.6	**21**	**16.9**
	D	4	3.2	4	3.2	3	2.4	0	0.0	3	2.4	**14**	**11.3**
	E	7	5.6	6	4.8	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	2	1.6	15	12.1	3	2.4	0	0.0	3	2.4	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

The results show that 23(18.6%), 8(6.4%), 9(7.2%), 8(6.4%), 12(9.6%) and 13(10.5%) of the respondents from research institute A, B, C, D, E and F respectively stated that the institute generates approximately 1-500GB of research data in every research project. In addition, respondents 21(17.0%), 7(5.6%), 12(9.7%) and 8(6.4%) from research institute A, B, C, and D in that order pointed out that the institute generated data approximately 500-1000GB. Approximately 1-500TB of research data generated in every research project was indicated by 20(16.2%), 6(4.8%) and 12(9.7%) respondents from research institutes A, B and C correspondingly. Further, the same number of respondents 7(5.6%), 13(10.5%) and 17(13.7%) from research institutes D, E and F respectively generated approximately 500-1000TB of research data in every research project. Furthermore, 20(16.2%), 11(8.9%), and 17(13.7%) of the respondents from research institutes A, C, and F correspondingly agreed that their institutes had generated 1-500PB and in the same group of respondents, 2(1.6%), 2(1.6%) and 3(2.4%) disagreed that they did not generate data amounting to 1-500PB. Generating data amounting >500 PB was indicated by 8(6.4%) and 17(13.7%) respondents from research institute D and F and on the other hand, the same group of respondents 3(2.4%) were undecided on generating research data more than 500PB.

5.5.2 Appraisal of research data

The respondents were also asked to state the research data appraisal checklist in their institute. The findings are presented in Table 5.7.

Table 5.7: Appraisal of research data (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	F	%	f	%	f	%
Uniqueness	28	22.6	64	51.6	10	8.1	21	16.9	1	0.8	124	100.0
Repeatability	36	29.0	79	63.7	7	5.6	2	1.6	0	0.0	124	100.0
Science/historical value	43	34.7	67	54.0	8	6.5	6	4.8	0	0.0	124	100.0
Complementary/added value	42	33.9	62	50.0	9	7.3	11	8.9	0	0.0	124	100.0
Reuse value	50	40.3	57	46.0	17	13.7	0	0.0	0	0.0	124	100.0
Substantiveness	32	25.8	75	60.5	17	13.7	0	0.0	0	0.0	124	100.0
Access	36	29.0	59	47.6	24	19.4	3	2.4	2	1.6	124	100.0
Volume	25	20.2	56	45.2	25	20.2	17	13.7	1	0.8	124	100.0
Cost-effectiveness	22	17.7	74	59.7	18	14.5	8	6.5	2	1.6	124	100.0

Source: Field data (2017)

The findings indicate that 92(74.2%) of the respondents agreed that uniqueness was one of the items considered in appraisal of research data whereas 22(17.7%) of the respondents disagreed that they did not appraise research data using uniqueness as a checklist. Further, 115(92.7%), 110(88.7%), 104(83.9%) and 107(86.3%) of the respondents respectively stated that research data appraisal was based on repeatability, scientific/historical value, complementary/added valued and reuse value. In addition, 107(86.3%) of the respondents stated that substantiveness was another appraisal checklist. The study established that 95(76.6%), 81(65.4%) and 96(77.4%) of the respondents stated that access, volume and cost-effectiveness were the items considered in data appraisal checklist in their institute. Nevertheless, 18(14.5%) and 10(8.1%) disagreed that they did not appraise research data using volume and cost-effectiveness as checklist.

5.5.2.1 Tools used to guide the appraisal of research data

The responses on tools used to guide the appraisal of research data are presented in Table 5.8.

Table 5.8 Cross tabulation of agricultural research institutes and tools used to guide the appraisal of research data (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Appraisal and selection policy	**A**	8	6.5	9	7.3	9	7.3	8	6.5	1	0.8	**35**	**28.2**
	B	7	5.6	4	3.2	3	2.4	2	1.6	0	0.0	**16**	**12.9**
	C	3	2.4	13	10.5	3	2.4	1	0.8	1	0.8	**21**	**16.9**
	D	1	0.8	6	4.8	5	4.0	2	1.6	0	0.0	**14**	**11.3**
	E	0	0.0	6	4.8	5	4.0	3	2.4	1	0.8	**15**	**12.1**
	F	0	0.0	6	4.8	12	9.7	5	4.0	0	0.0	**23**	**18.5**
RDM policy	**A**	5	4.0	12	9.7	14	11.3	4	3.2	0	0.0	**35**	**28.2**
	B	2	1.6	6	4.8	7	5.6	1	0.8	0	0.0	**16**	**12.9**
	C	1	0.8	14	11.3	3	2.4	3	2.4	0	0.0	**21**	**16.9**
	D	3	2.4	4	3.2	4	4.2	3	2.4	0	0.0	**14**	**11.3**
	E	2	1.6	9	7.3	3	2.4	1	0.8	0	0.0	**15**	**12.1**
	F	5	4.0	2	1.6	13	10.5	3	2.4	0	0.0	**23**	**18.5**
Research institute policy	**A**	9	7.3	12	9.7	10	8.1	4	3.2	0	0.0	**35**	**28.2**
	B	4	3.2	7	5.6	4	3.2	1	0.8	0	0.0	**16**	**12.9**
	C	5	4.0	11	8.9	5	4.0	0	0.0	0	0.0	**21**	**16.9**
	D	3	2.4.	1	0.8	6	4.8	4	3.2	0	0.0	**14**	**11.3**
	E	3	2.4	10	8.1	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	8	6.5	4	3.2	11	8.9	0	0.0	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

It is apparent from this Table 5.8 that 17(13.8%), 11(8.8%) and 16(12.9%) of the respondents from research institute A, B and C respectively stated that appraisal and selection policy was one of the tools used to guide the appraisal of research data whereas 2(1.6%) of the respondents each from research institute B, C and D disagreed that they did not appraisal and selection policy was used in appraisal of research data. There were 7(5.6%), 11(8.9%) and 7(5.6%) of the respondents from research institutes D, E and F who stated that RDM policy was used to guide the appraisal of research data and on the contrary 3(2.4%) of the respondent each from research institutes C, D and F disagreed that they did not use RDM policy in appraising research data. Another 21(17.0%), 16(12.9%) and 13(10.5%) of the respondents from research institutes A, C and E correspondingly asserted that research institute's policy was used to guide the appraisal of research data whereas 4(3.2%), 1(0.8%) and 4(3.2%) of the respondents from research institutes A, B and D in that order disagreed that they did not use research institute policy to appraise research data. One head of research [HR6] interviewed from research institute F concurred with the researchers and stated:

> *"Data Appraisal is done where raw data is recorded or collected in notebooks, data sheets or in a computer then subjected to section according to its relevance to the purpose. In the selection process researchers are supposed to follow SOPs and the institute's policy to determine what data to keep"*

5.5.3 Description of research data (metadata)

There was need to establish the respondent's opinions concerning the description of research data. The respondents were asked to describe each of their research datasets. The responses are presented in Table 5.9.

Table 5.9 Cross tabulation of agricultural research institutes and description of research datasets (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
The only description will be the filename on my hard drive	**A**	7	5.6	14	11.3	10	8.1	3	2.4	1	0.8	**35**	**28.2**
	B	2	1.6	7	5.6	5	4.0	1	0.8	1	0.8	**16**	**12.9**
	C	4	3.2	15	12.1	1	0.8	1	0.8	0	0.0	**21**	**16.9**
	D	1	0.8	7	5.6	4	3.2	2	1.6	0	0.0	**14**	**11.3**
	E	5	4.0	8	6.5	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	5	4.0	12	9.7	6	4.8	0	0.0	0	0.0	**23**	**18.5**
I will use on my handwritten notes lab notebook	**A**	7	5.6	14	11.3	10	8.1	3	2.4	1	0.8	**35**	**28.2**
	B	2	1.6	6	4.8	6	4.8	1	0.8	1	0.8	**16**	**12.9**
	C	6	4.8	13	10.5	2	1.6	0	0.0	0	0.0	**21**	**16.9**
	D	2	1.6	7	5.6	4	3.2	1	0.8	0	0.0	**14**	**11.3**
	E	6	4.8	9	7.3	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	0	0.0	20	16.1	3	2.4	0	0.0	0	0.0	**23**	**18.5**
I will describe the data using the column and row labels	**A**	5	4.0	20	16.1	8	6.5	2	1.6	0	0.0	**35**	**28.2**
	B	2	1.6	9	7.3	5	4.0	0	0.0	0	0.0	**16**	**12.9**
	C	4	3.2	14	11.3	2	1.6	1	0.8	0	0.0	**21**	**16.9**
	D	2	1.6	8	6.5	2	1.6	2	1.6	0	0.0	**14**	**11.3**

	E	4	3.2	10	8.1	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	5	4.0	13	10.5	5	4.0	0	0.0	0	0.0	**23**	**18.5**
I will create descriptive metadata	**A**	4	3.2	16	12.9	7	5.6	4	3.2	4	3.2	**35**	**28.2**
	B	2	1.6	8	6.5	4	3.2	2	1.6	0	0.0	**16**	**12.9**
	C	6	4.8	12	9.7	0	0.0	2	1.6	1	0.8	**21**	**16.9**
	D	4	3.2	7	5.6	2	1.6	1	0.8	0	0.0	**14**	**11.3**
	E	3	2.4	7	5.6	1	0.8	1	0.8	3	2.4	**15**	**12.1**
	F	0	0.0	15	12.1	3	2.4	0	0.0	5	4.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes**: A, B, C, D, E, and F

Cross tabulation of agricultural research institutes and description of research data was done. The findings in Table 5.9 show that 21(16.9%), 9(7.2%) and 19(15.3%) of the respondents from research institutes A, B, and C respectively stated that the only description will be the filenames on their hard drive. However, 4(3.2%), 2(1.6%) and 2(1.6%) of the respondents from research institutes A, B and D correspondingly disagreed that the filenames on their hard drive was not the only description used. Similarly, 21(16.9%), 19(15.3%) and 20(16.1%) of the respondents from research institutes A, C, and F respectively stated that they use handwritten notes on their lab notebook after the experiments had been completed. On the contrary, 4(3.2%), 2(1.6%) and 1(0.8%) of the respondent from research institutes A, B and D disagreed that they did not use handwritten notes on their lab notebook. The results further show that research institutes A, B, C and D showed that 25(20.2), 11(8.9%), 18(14.5% and 10(8.1%) of the respondents in that order agreed that they had described data using the column and row labels in their spreadsheets after the data was analyzed. The study also established that 20(16.1%), 18(14.5%) and 11(8.9%) of the participants from research institute A, C and D correspondingly asserted that they create descriptive metadata for each dataset and save the descriptions with datasets on hard drive. On the contrary, 8(6.4%) and 5(4.0%) respondent from research institutes A and F disagreed that they did not create descriptive metadata for each dataset and save the descriptions with datasets on hard drive.

5.5.3.1 Creating or capturing metadata

Properly describing and documenting research data allows users to understand and track important details of their work. In addition, having metadata about the research data also facilitates search and retrieval because well described metadata records make research data collections discoverable, citable, reusable and accessible for the long term in data repository (Australian National Data Service, n.d (b)). Besides, metadata includes content such as contact information, geographic locations, details about units of measure, abbreviation or codes used in dataset, survey tools details and much more.

Figure 5.2 shows results on creating or capturing of metadata in agricultural research institutes.

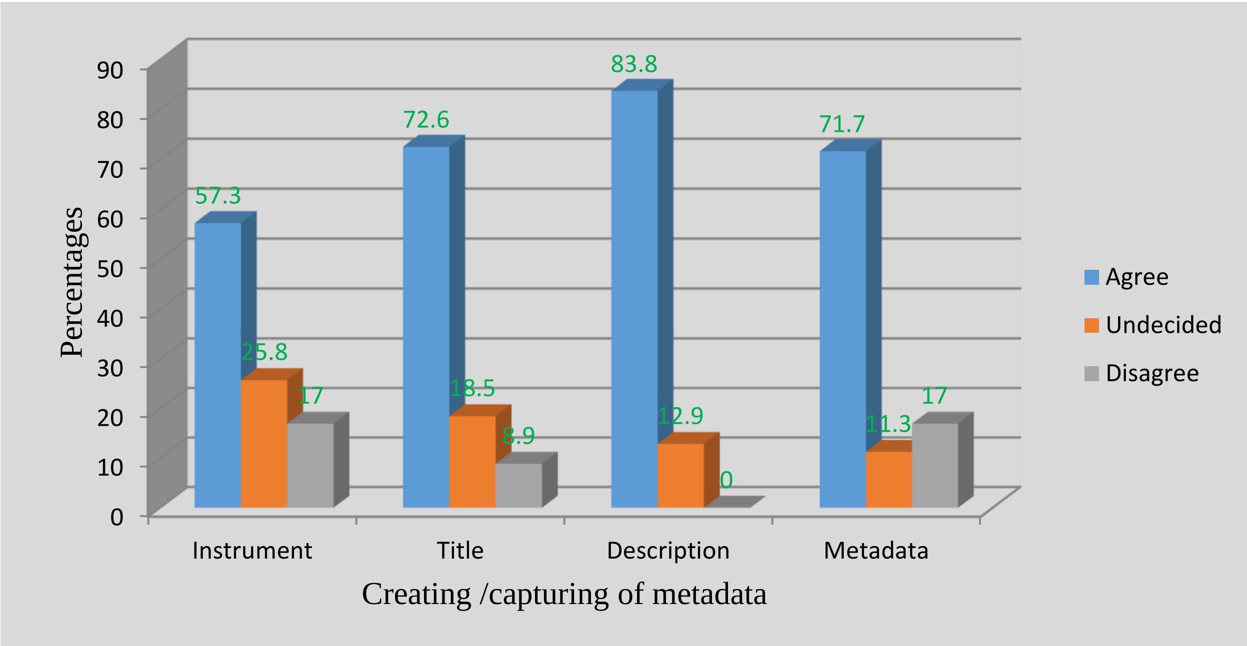

Figure 5.2 Creating/ capturing of metadata (n=124)

Concerning the creation and capturing of metadata, the findings indicate that 71(57.3%) of the respondents stated that instrument metadata are automatically included in each data file, while 21(17%) disagreed that instrument metadata are not automatically included in each data file. There were 90(72.6%) of the respondents who stated that they do create a title and short textual description for each dataset when submitting the dataset to their research institute data repository. Another 104(83.9%) agreed that data descriptions were be saved in spreadsheets or word processor documents while 20(16.1%) of the respondents disagreed that they their data description are not saved in spreadsheets or processor documents. Finally, majority 89(71.8%) of the respondents stated that they do create rich metadata by recording data at the time of capturing using a metadata entry form to ensure they don't miss any essential information. The respondents stated that the metadata file will be saved locally with their dataset, and eventually deposited with the dataset when it is submitted to a data repository. However, 21(17%) of the respondents had an opposing opinion, disagreeing that they did not create rich metadata by recording data at the time of capturing using a metadata entry form.

5.5.4 Storage, backups and preservations of research data

Research data storage, backup and preservation is important (University of Oregon Libraries, n.d) because data may need to be accessed in the future to explain or augment subsequent research, safeguard research investments; other researchers might wish to evaluate or use the results of other researchers; and stored data can establish precedence in the event that similar research is published. In essence, data storage and preservation are key elements in the research data lifecycle.

5.5.4.1 Storage of research data

Data storage is a key element in the research data lifecycle and for this reason it is important to think at the beginning of research project how and where research data will be stored. The storage environment should be actively managed, secure and reliable over time and should ensure that it enables the level of control and accessibility required by the researcher(s) and others who require access, use and reuse of data. The section will address questions on storage of research data.

5.5.4.1.1 Storage of research data in short term after acquisition

The respondents were asked to state where they stored their research data in the short term, after acquisitions. The responses are presented in Table 5.10.

Table 5.10 Cross tabulation of agricultural research institutes and storage of research data in short term, after acquisition (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
On my laptop	**A**	18	14.5	10	8.1	4	3.2	2	1.6	1	0.8	**35**	**28.2**
	B	6	4.8	5	4.0	3	2.4	1	0.8	1	0.8	**16**	**12.9**
	C	13	10.5	8	6.5	0	0.0	0	0.0	0	0.0	**21**	**16.9**
	D	9	7.3	4	3.2	1	0.8	0	0.0	0	0,0	**14**	**11.3**
	E	7	5.6	7	5.6	0	0.0	1	0.8	0	0.0	**15**	**12.1**
	F	18	14.5	0	0.0	0	0.0	3	2.4	2	1.6	**23**	**18.5**
On the computer	**A**	9	7.3	9	7.3	6	4.8	7	5.6	4	3.2	**35**	**28.2**
Connected to the serve	**B**	4	3.2	5	4.0	3	2.4	1	0.8	3	2.4	**16**	**12.9**
	C	5	4.0	10	8.1	2	1.6	3	2.4	1	0.8	**21**	**16.9**
	D	4	3.2	4	3.2	3	2.4	2	1.6	1	0.8	**14**	**11.3**
	E	4	3.2	4	3.2	2	1.6	4	3.2	1	0.8	**15**	**12.1**

	F	0	0.0	2	1.6	4	3.2	12	9.7	5	4.0	**23**	**18.5**
On my research group's data storage file	**A**	7	5.6	17	13.7	6	4.8	2	1.6	3	2.4	**35**	**28.2**
	B	4	3.2	9	7.3	1	0.8	1	0.8	1	0.8	**16**	**12.9**
	C	4	3.2	9	7.3	5	4.0	1	0.8	2	1.6	**21**	**16.9**
	D	3	2.4	4	3.2	4	3.2	1	0.8	2	1.6	**14**	**11.3**
	E	0	0.0	11	8.9	1	0.8	1	0.8	2	1.6	**15**	**12.1**
	F	3	2.4	20	16.1	0	0.0	0	0.0	0	0.0	**23**	**18.5**
CDs or DVDs	**A**	8	6.5	13	10.5	9	7.3	5	4.0	0	0.0	**35**	**28.2**
	B	2	1.6	4	3.2	6	4.8	4	3.2	0	0.0	**16**	**12.9**
	C	2	1.6	15	12.1	4	3.2	0	0.0	0	0.0	**21**	**16.9**
	D	6	4.8	3	2.4	3	2.4	2	1.6	0	0.0	**14**	**11.3**
	E	4	3.2	9	7.3	1	0.8	1	0.8	0	0.0	**15**	**12.1**
	F	0	0.0	18	14.5	2	1.6	3	2.4	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

The findings indicate that 28(22.6%), 11(8.8%0), 21(17.0%) of the respondents from research institutes A, B and C respectively stated that they stored the research data on their laptops. Likewise, 13(10.5%), 14(11.2%), 18(14.5%) of the respondents from research institutes D, E and F in that order asserted that they stored the research data on their laptops while there was no respondents 0(0.0%) who disagreed from research institutes C and D that they do not use their laptops to store research data. On storing their research data on computer connected to the server, 8(6.4%), 8(6.4%) and 2(1.6%) of the respondents from research institutes D, E and F correspondingly agreed that they do. On the contrary, 11(8.8%), 4(3.2%) and 17(13.7%) of the respondents from research institutes A, C and F in that order disagreed that they did not store research data in the short term on the computer connected to the server. Further in research institutes A, B and C, respondents 24(19.3%), 13(10.5%) and 13(10.5%), respectively stated that they stored their research data on their research group's data storage file whereas 21(17.0%), 6(4.8%), 17(13.7) respondents from research institutes A, B and C respectively agreed that they stored their research data on CDs or DVDs. However, 9(7.3%) and 6(4.8%) of the respondents each from research institutes A and B was undecided on storing their research data on CDs or DVDs while 5(4.0%) and 4(3.2%) of the same group of respondents disagreed that they did not store their data on CD or DVDs.

5.5.4.1.2 Storage media of research data preservation

The respondents were also asked to state the storage media of their research data awaiting preservation. The responses are presented in figure 5.3.

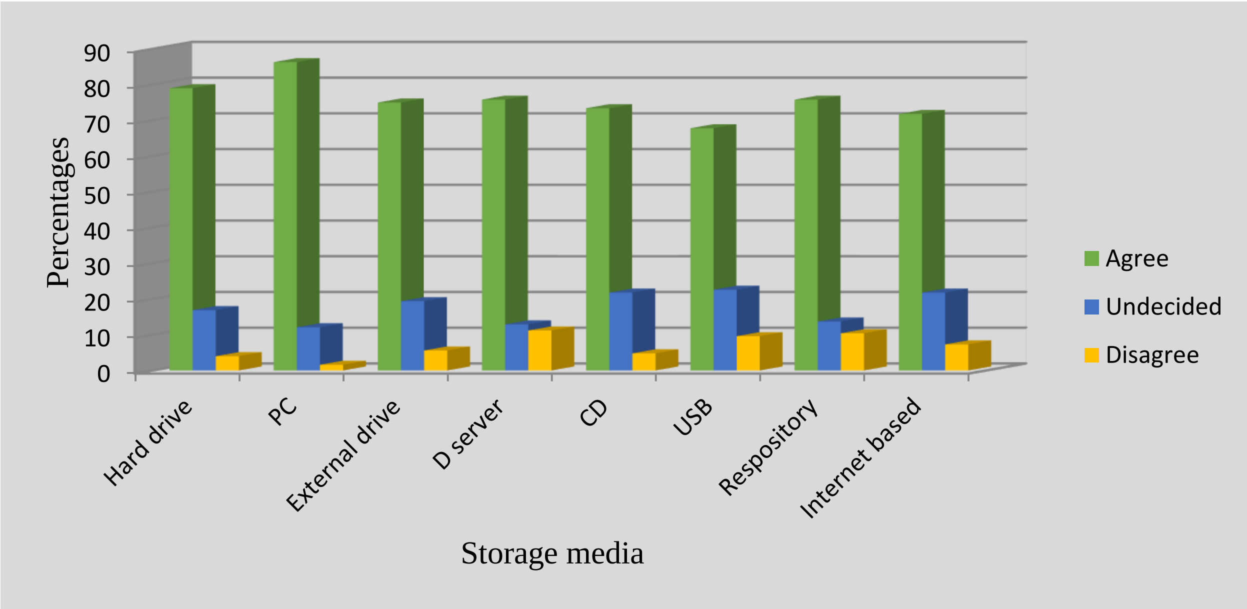

Figure 5.3 Storage media awaiting preservation (n=124)

The study revealed that 98(79%) of the respondents store research data in the hard drive of the instruments which generate the data, further 107(86.3%), 93(75%) and 94(75.8%) of the respondents stated that the research data was stored in PC hard drive, external hard drive and departmental servers respectively. There were 91(73.4%), 84(67.7%) and 94(75.8%) of the respondents who stated that they stored their research data in CD/DVD, USB flash drive and institutional repository respectively. However, 13(10.4%) disagreed that they did not store their research data in the institutional repository. Another 89(71.8%) agreed that they used internet-based storage.

5.5.4.2 Backups of research data

Backing up research data is essential to avoid the risk of losing data through accidental deletion, hard-drive failure, or theft or damage of equipment. Backup strategies should begin at the start of research because research data generated are of high value, unique or less easily reproduced as a result safeguarding research datasets through storage and backup are essential. This section will address questions on backup of research data.

5.5.4.2.1 Responsibility for day-to-day management, storage and backup of research data

The study sought to determine the person responsible for the day-to-day management, storage and backup of the data arising from the research. The findings are shown in Table 5.11.

Table 5.11 Cross tabulation of agricultural research institutes and responsibility for day-to-day management, storage and backup of research data (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Myself	**A**	14	11.3	8	6.5	5	4.0	6	4.8	2	1.6	**35**	**28.2**
	B	7	5.6	2	1.6	2	1.6	3	2.4	2	1.6	**16**	**12.9**
	C	5	4.0	10	8.1	4	3.2	2	1.6	0	0.0	**21**	**16.9**
	D	4	3.2	4	3.2	2	1.6	3	2.4	1	0.8	**14**	**11.3**
	E	5	4.0	7	5.6	2	1.6	0	0.0	1	0.8	**15**	**12.1**
	F	16	12.9	7	5.6	0	0.0	0	0.0	0	0.0	**23**	**18.5**
Group Data	**A**	15	12.1	10	8.1	8	6.5	0	0.0	2	1.6	**35**	**28.2**
manager	**B**	4	3.2	4	3.2	6	4.8	0	0.0	2	1.6	**16**	**12.9**
	C	6	4.8	11	8.9	3	2.4	1	1.6	0	0.0	**21**	**16.9**
	D	5	4.0	5	4.0	3	2.4	0	0.0	1	0.8	**14**	**11.3**
	E	4	3.2	9	7.3	1	0.8	1	1.6	0	0.0	**15**	**12.1**
	F	11	8.9	10	8.1	2	1.6	0	0.0	0	0.0	**23**	**18.5**
Department	**A**	4	3.2	18	14.5	7	5.6	2	1.6	4	3.2	**35**	**28.2**
	B	2	1.6	7	5.6	4	3.2	1	0.8	2	1.6	**16**	**12.9**
	C	6	4.8	9	7.3	3	2.4	3	2.4	0	0.0	**21**	**16.9**
	D	4	3.2	4	3.2	4	3.2	1	0.8	1	0.8	**14**	**11.3**
	E	1	0.8	10	8.1	1	0.8	1	0.8	2	1.6	**15**	**12.1**
	F	8	6.5	10	8.1	2	1.6	0	0.0	3	2.4	**23**	**18.5**
IT staff	**A**	3	2.4	14	11.3	9	7.3	3	2.4	6	4.8	**35**	**28.2**
	B	1	0.8	8	6.5	3	2.4	2	1.6	2	1.6	**16**	**12.9**
	C	2	1.6	10	8.1	6	4.8	1	0.8	2	1.6	**21**	**16.9**
	D	3	2.4	2	1.6	6	4.8	1	0.8	2	1.6	**14**	**11.3**
	E	1	0.8	6	4.8	4	3.2	0	0.0	4	3.2	**15**	**12.1**
	F	0	0.0	13	10.5	5	4.0	0	0.0	5	4.0	**23**	**18.5**

Librarian	**A**	3	2.4	10	8.1	12	9.7	3	2.4	7	5.6	**35**	**28.2**
	B	2	1.6	4	3.2	9	7.3	0	0.0	1	0.8	**16**	**12.9**
	C	1	0.8	9	7.3	6	4.8	2	1.8	3	2.4	**21**	**16.9**
	D	1	0.8	3	2.4	3	2.4	3	2.4	4	3.2	**14**	**11.3**
	E	2	1.6	4	3.2	4	3.2	0	0.0	5	4.0	**15**	**12.1**
	F	2	1.6	3	2.4	0	0.0	8	6.5	10	8.1	**23**	**18.5**
The research unit	**A**	10	8.1	9	7.3	6	4.8	5	4.0	5	4.0	**35**	**28.2**
	B	8	6.5	4	3.2	2	1.6	0	0.0	2	1.6	**16**	**12.9**
	C	4	3.2	9	7.3	4	3.2	2	1.6	2	1.6	**21**	**16.9**
	D	3	2.4	3	2.4	4	3.2	2	1.6	2	1.6	**14**	**11.3**
	E	1	0.8	7	5.6	3	2.4	2	1.6	2	1.6	**15**	**12.1**
	F	2	1.6	9	7.3	4	3.2	5	4.0	3	2.4	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

As shown in Table 5.11, a significant proportion of respondents 22(17.8%), 15(12.1%) and 23(18.5%) from research institutes A, C and F correspondingly stated that researchers (myself) are responsible for day-to-day management, storage and backup of the data. Furthermore, 8(6.4%), 10(8.0%) and 13(10.5%) of the respondents from research institutes B, D and E respectively stated that the research group's data manager was responsible for day-to-day management of the research data. According to 22(17.7%), 9(7.2%), 15(12.1%), 8(6.4%), 11(8.9%) and 18(14.6%) of the respondents from research institutes A, B, C, D, E and F respectively indicated that the department was responsible for day-to-day management, storage and backup of the data arising from the research whereas 3(2.4%) of the respondents each from research institutes B and E disagreed that the department was not responsible for day-to-day management, storage and backup of research data . Further, most respondents 17(13.7%), 12(9.7%) and (13(10.5%) of the respondents from research institutes A, C and F in that order stated that IT staff were responsible for day-to-day management, storage and backup of the research data because they use cloud storage whereas 6(4.8%) of the respondents each from research institutes C and D were undecided that IT staff were responsible. Another 13(10.5%), 6(4.8%) and 4(3.2%) from research institutes A, B and D in that order stated that data management was done by librarian. However 1(0.8%) and 7(5.6%) of the respondents from research institutes B and D respectively disagreed that day-to-day management, storage and backup of the research data was not the responsibility of the librarian.

The results also show that 19(15.4%), 13(10.5%) and 8(6.4%) of the respondents from research institutes A, C and E correspondingly stated that the day-to-day management, storage and backup of the data was done at the research unit while 4(3.2%) of the respondents each from research institutes C, D, E disagreed that the responsibility of data management, storage and backup did not rest on the research unit.

Besides, interviews conducted on directors and heads of research concurred with researchers that the heads of department were responsible for day-to-day management, storage and backup of the research data generated in their departments

5.5.4.2.2 Frequency of making backups for the research data

The respondents were also asked to state the frequency of making backups for the research data. The results are shown in Table 5.12.

Table 5.12 Frequency of making backups (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	F	%	f	%
Hourly	20	16.1	51	41.1	20	16.1	20	16.1	13	10.5	124	100.0
Daily	28	22.6	61	49.2	20	16.1	11	8.9	4	3.2	124	100.0
Weekly	27	21.8	58	46.8	19	15.3	13	10.5	7	5.6	124	100.0
Monthly	23	18.5	58	46.8	20	16.1	14	11.3	9	7.3	124	100.0
Annually	15	12.1	38	30.6	25	20.2	18	14.5	28	22.6	124	100.0
Never	25	20.2	45	36.3	19	15.3	16	12.9	19	15.3	124	100.0

Source: Field data (2017)

The results revealed that 71(57.2%) of the respondents stated that they made backups hourly while 89(71.8%) made backups daily and more than half 85(68.6%) made backups weekly. The results show that 81(65.3%) of the respondents made backups monthly whereas less than half 53(42.7%) made backups annually. However, 70(56.5%) of the respondents stated that they had never made backups and 35(28.2%) disagreed that they did not make backups.

5.5.4.3 Preservation of research data

A number of preservation actions are required before research data is integrated into the curation environment to ensure that their authoritative characteristics, as identified by ISO 15489-1, can be retained for long term (Higgins, 2012). There are a range of actions associated with data preservation for instance preserving and sharing, selecting data for long term curation, choosing a data service for data, licenses for research data, discoverability and data access statements. This section will address questions on preservation of research data.

5.5.4.3.1 Length of time for keeping research data

The study also sought to know the length of time for keeping research data in the institution before it was disposed. The results are presented in Figure 5.4.

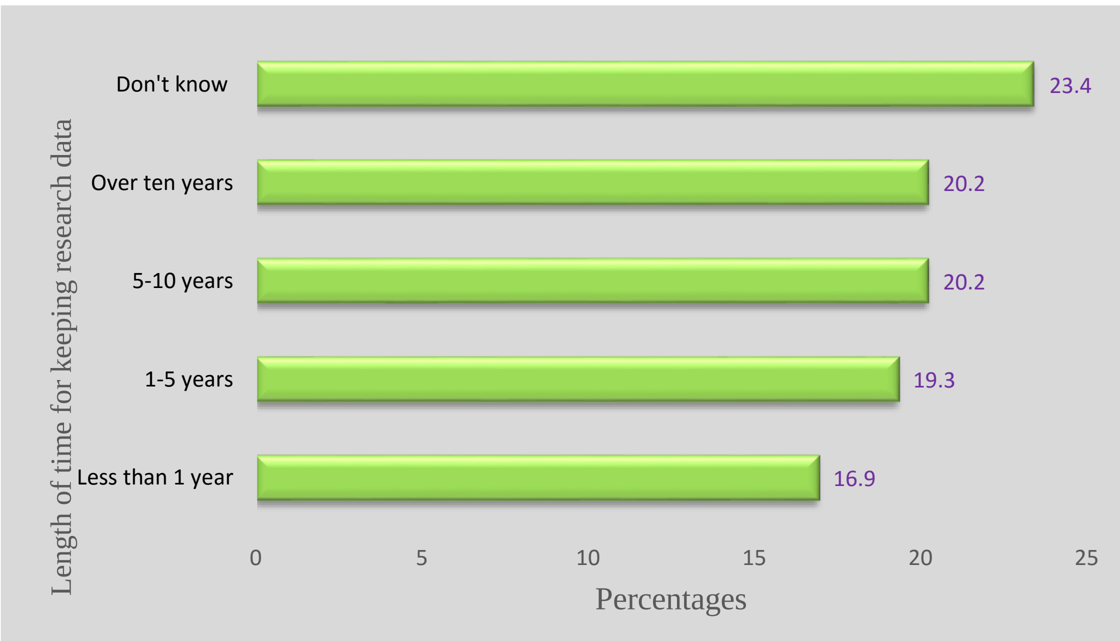

Figure 5.4 Length of time for keeping research data (n=124)

The results showed that 21(16.9%) and 24(19.3%) of the respondents stated that the research data was kept for a period of less than one year and 1-5 years respectively before it was disposed. Another 25(20.2%) stated that the research data was stored in their institutions for a period of 5-10 years whereas 25(20.2%) agreed that data was stored for more than 10 years. Further, 29(23.4%) of the respondents didn't know how long data was kept in their research institute before it was disposed. Through interviews, one institute director [D3] from Research institute C had this to say:

> *"Keeping of research is indefinite since it builds up on what researchers are doing. Data is rarely destroyed and helps in checking trends. There is data as far as 1907 and it is still useful. In agricultural research, data can be referred to, use same data to make new insights in other areas and also observe trends. We also have a minimum period that data can be preserved but that does not mean when that time elapses, we destroy. No, we do not".*

The findings reveal that preserving agricultural research data is important, especially if it is guided by preservation policy.

5.5.4.3.2 Research data worth preserving

The study wanted to determine whose responsibility it was to decide which research data were worth preserving. The results are presented in Table 5.13.

Table 5.13 Cross tabulation of agricultural research institutes and responsibility for deciding research data worth preserving (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Myself alone	**A**	16	12.9	7	5.6	2	1.6	2	1.6	8	6.5	**35**	**28.2**
	B	6	4.8	0	0.0	2	1.6	0	0.0	8	6.5	**16**	**12.9**
	C	13	10.5	7	5.6	0	0.0	0	0.0	1	0.8	**21**	**16.9**
	D	6	4.8	5	4.0	1	0.8	2	1.6	0	0.0	**14**	**11.3**
	E	4	3.2	11	8.9	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	12	9.7	11	8.9	0	0.0	0	0.0	0	0.0	**23**	**18.5**
Myself/research supervisor	**A**	15	12.1	10	8.1	4	3.2	3	2.4	3	2.4	**35**	**28.2**
	B	10	8.1	0	0.0	1	0.8	3	2.4	2	1.6	**16**	**12.9**
	C	14	11.3	6	4.8	1	0.8	0	0.0	0	0.0	**21**	**16.9**
	D	3	2.4	8	6.5	2	1.6	0	0.0	1	0.8	**14**	**11.3**
	E	7	5.6	8	6.5	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	12	9.7	8	6.5	0	0.0	3	2.4	0	0.0	**23**	**18.5**
My research	**A**	8	6.5	6	4.8	9	7.3	5	4.0	7	5.6	**35**	**28.2**

supervisor alone	**B**	4	3.2	0	0.0	4	3.2	2	1.6	6	4.8	**16**	**12.9**
	C	5	4.0	7	5.6	3	2.4	5	4.0	1	0.8	**21**	**16.9**
	D	2	1.6	4	3.2	5	4.0	2	1.6	1	0.8	**14**	**11.3**
	E	2	1.6	4	3.2	4	3.2	3	2.4	2	1.6	**15**	**12.1**
	F	3	2.4	18	14.5	0	0.0	2	1.6	0	0.0	**23**	**18.5**
Institutional policy	**A**	12	9.7	11	8.9	7	5.6	1	0.8	4	3.2	**35**	**28.2**
	B	9	7.3	0	0.0	2	1.6	1	0.8	4	3.2	**16**	**12.9**
	C	9	7.3	9	7.3	1	0.8	2	1.6	0	0.0	**21**	**16.9**
	D	2	1.6	8	6.5	2	1.6	1	0.8	1	0.8	**14**	**11.3**
	E	1	0.8	11	8.9	3	2.4	0	0.0	0	0.0	**15**	**12.1**
	F	3	2.4	17	13.7	0	0.0	3	2.4	0	0.0	**23**	**18.5**
Funders requirement	**A**	11	8.9	14	11.3	7	5.6	3	2.4	0	0.0	**35**	**28.2**
	B	7	5.6	5	4.0	3	2.4	1	0.8	0	0.0	**16**	**12.9**
	C	3	2.4	14	11.3	2	1.6	2	1.6	0	0.0	**21**	**16.9**
	D	4	3.2	4	3.2	4	3.2	2	1.6	0	0.0	**14**	**11.3**
	E	1	0.8	14	11.3	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	16	12.9	7	5.6	0	0.0	0	0.0	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

Majority 23(18.6%), 20(16.1%) and 23(18.6%) of the respondents from research institutes A, C and F respectively stated that researchers (myself alone) decided which research data was worth preserving while 8(6.4%) and 2(1.6%) from research institutes B and D in that order disagreed that researchers were not responsible for deciding which research data was worth preserving. Similarly, 20(16.1%) of the respondents each from research institutes C and F agreed that researchers (myself) decided on the worth of research data to be preserved in consultation with their research supervisor. Further still, a significant proposition of respondents 14(11.3%), 12(9.6%) and 21(16.9%) from research institutes A, C and F respectively indicated that the decision on the worthiness of research data to be preserved was done by the research supervisor alone, however 12(9.7%), 6(4.8%) and 2(1.6%) of the respondents from the same research institutes disagreed that the research supervisor was not the only one deciding the worthiness of research data to be preserved. It was found that institutional policy was used to determine the research data that were worth preserving as stated by 9(7.3%), 18(14.6%) and 12(9.7%) of the respondents from research institutes B, C and E correspondingly. The results in Table 5.13 also show that most respondents 25(20.2%) and 23(18.5%) from research institutes A and F respectively pointed out that research data to be preserved was decided by funders requirement whereas 3(2.4%) and 0(0.0%) from the same research institutes disagreed that research data worth preserving was not decided by funders requirement.

5.5.4.3.3 When research data is moved to a secure archive for long-term preservation

The respondents were also asked to state the time when their research data was moved to a secure archive for long-term preservation and publication. The responses are presented in Table 5.14.

Table 5.14 When research data is moved to a secure archive for long-term preservation (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Upon completion of each set of experiments	23	18.5	58	46.8	20	16.1	15	12.1	8	6.5	124	100.0
When my research group leader decides	12	9.7	54	43.5	25	20.2	12	9.7	21	16.9	124	100.0
Immediately after publication of my paper	23	18.5	70	56.5	16	12.9	5	4.0	10	8.1	124	100.0
Determined by RDM policy	33	26.6	55	44.4	10	8.1	13	10.5	13	10.5	124	100.0

Source: Field data 2017

As shown in Table 5.14, data was moved to a secure archive for long-term preservations upon completion of each set experiments as agreed by majority of the respondents 81(65.3%) and when their research group leader decides it is appropriate as stated by 66(53.2%) of respondents . The research data was also moved to a secure archive for long term preservation immediately after publication of the paper in a journal. This was according to 93(75%) of the respondents. Further 88(71%) of the respondents stated that the time for moving research data to a secure archive for long-term preservations was determined by RDM policy.

5.5.4.3.4 Location for archiving research data for long-term preservation

This study also wanted to know where the research data was archived for long-term preservation. The responses are presented in Table 5.15.

Table 5.15 Cross tabulation of agricultural research institutes and location for archiving research data for long-term preservation (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Selected data will be included in the figures and tables of research papers published	**A**	6	4.8	20	16.1	4	3.2	2	1.6	3	2.4	**35**	**28.2**
	B	2	1.6	11	8.9	0	0.0	0	0.0	3	2.4	**16**	**12.9**
	C	8	6.5	7	5.6	2	1.6	3	2.4	1	0.8	**21**	**16.9**
	D	0	0.0	7	5.6	4	3.2	2	1.6	1	0.8	**14**	**11.3**
	E	4	3.2	10	8.1	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	3	2.4	17	13.7	0	0.0	0	0.0	3	2.4	**23**	**18.5**
As supplementary files attached to my journal articles on the Publisher's web site.	**A**	11	8.9	16	12.9	6	4.8	2	1.6	0	0.0	**35**	**28.2**
	B	4	3.2	11	8.9	1	0.8	0	0.0	0	0.0	**16**	**12.9**
	C	4	3.2	13	10.5	2	2.4	1	0.8	1	0.8	**21**	**16.9**
	D	3	2.4	3	2.4	5	4.0	2	1.6	1	0.8	**14**	**11.3**
	E	2	1.6	10	8.1	1	0.8	2	1.6	0	0.0	**15**	**12.1**
	F	3	2.4	18	14.5	0	0.0	2	1.6	0	0.0	**23**	**18.5**
In the research institute's data Archive server	**A**	9	7.3	15	12.1	9	7.3	2	1.6	0	0.0	**35**	**28.2**
	B	5	4.0	7	5.6	4	3.2	0	0.0	0	0.0	**16**	**12.9**
	C	7	5.6	11	8.9	1	0.8	0	0.0	2	1.6	**21**	**16.9**
	D	2	1.6	5	4.0	5	4.0	1	0.8	1	0.8	**14**	**11.3**
	E	2	1.6	9	7.3	4	3.2	0	0.0	0	0.0	**15**	**12.1**
	F	6	4.8	17	13.7	0	0.0	0	0.0	0	0.0	**23**	**18.5**
In the research institute's Data Bank	**A**	11	8.9	11	8.9	9	7.3	4	3.2	0	0.0	**35**	**28.2**
	B	5	4.0	3	2.4	5	4.0	3	2.4	0	0.0	**16**	**12.9**
	C	6	4.8	13	10.5	0	0.0	1	0.8	1	0.8	**21**	**16.9**
	D	2	1.6	6	4.8	4	3.2	1	0.8	1	0.8	**14**	**11.3**
	E	4	3.2	9	7.3	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	7	5.6	16	12.9	0	0.0	0	0.0	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F

It is apparent from this Table 5.15 that a significant proportion of respondents 26(20.9%), 13(10.5%) and 20(16.1%) from research institutes A, B and F in that order stated that selected data were included in the figures and tables of the research papers published by their research groups. There were 17(13.7%), 6(4.8%) and 12(9.7%) of the respondents from research institutes C, D and E respectively who stated that the research data was stored as supplementary files attached to the journal articles on publisher's website. Another 24(19.4%), 18(14.5%) and 23(18.5%) of the respondents from research institutes A, C and F correspondingly asserted that the research data was archived in the research institute's data archive server. Similarly, 19(15.4%), 8(6.4%) and 13(10.5%) of the respondents from research institutes C, D and E correspondingly stated that the research data was archived in the research institute's Data Bank while, 2(1.6%) of the respondents each from research institute C and D disagreed that the location for archiving research data for long-term preservation was not the research institute's data bank.

Concerning preservation, majority of the respondents interviewed lamented that there is no clear policy and guidelines on where all research data generated should be preserved. However, each agricultural research institute had their own ways of preserving data either in personal lap-tops, external storage or with IT departments. Majority of the respondents were skeptical about preserving data on KALRO's server and repository. One head of research [HR15] had this to say:

> *'Data is always preserved in publications such as annual reports and journals such as Tea journals, bulletins, annual reports, pamphlets, cabinet storage, and an active website where there is a section on publication. Every agricultural research institute has their way of archiving research data for long term preservation. We do not have a policy on archiving as KALRO but I think there is need for one. It is necessary to have posterity of our research data. Otherwise we repackaged research data into information that the public can understand and make use of it.*

One head of IT [IT 1] pointed out:

> *"My IT department is not involve in preservation of research data instead it is involved in provision of infrastructure and software, creation and designing of front end forms to fill data, and trouble-shooting concerns while preservation is the preserve of researchers and heads of department. The most unfortunate thing is that research data are scattered all over in offices, researchers lap tops, cabinets in a very disorganized way……and with this kind of trend we will loss valuable data …so there is need to start RDM unit to collect all data and preserve"*

5.5.4.3.5 Means of transferring research datasets for long-term archiving

Further, the respondents were asked to state the means of transferring research data for long term archiving. Table 5.16 presents the results.

Table 5.16 Means of transferring research datasets for long-term archiving (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Using physical hard drives	33	26.6	49	39.5	25	20.2	14	11.3	3	2.4	124	100.0
By e-mailing files	20	16.1	47	37.9	38	30.6	11	8.9	8	6.5	124	100.0
By selected data repository's web-based	25	20.2	65	52.4	22	17.7	10	8.1	2	1.6	124	100.0
By use of a local data management system	21	16.9	57	46.0	24	19.4	9	7.3	13	10.5	124	100.0

Source: Field data 2017

The results revealed that 82(66.1%) of the respondents stated that the datasets were transferred on physical hard drives whereas 67(54%) stated that it was transferred by the e-mailing file to the IT department or librarian. Further, 90(72.6%) stated that the transfer of research data for long term archiving was by selected data repository's Web-based submission and 12(9.7%) of the respondents disagreed that transfer of research data for long term archiving was not by selected data repository's Web-based. There were 78(62.9%) of the respondents who stated that the research datasets were transferred using a local data management system such as DataStage that can automatically package and submit data files to the selected repository. However 24(19.4%) of respondents were undecided on the use of a local data management system whereas 22(17.8%) disagreed that they did not use a local data management system. One head of research [HR11] from Research institute E was interviewed and had this to say:

> *"There has been no agreement, policy or guidelines on the means of transferring data…… researchers have different types of data generated from different project, and the type of data will influence the means of transfer for long term preservation. To be sincere, the commonly used means of transfer is printing hard copy, external storage and emails. The issue of using a repository and serve is not accepted much by our researchers because of lack of trust and mainly issues of plagiarism……."*

5.5.4.4 Responsibility for RDM after leaving research institute

The respondents were also asked to state the person who was responsible for the research data management after researcher had left the research institute. The findings are presented in Table 5.17.

Table 5.17 Cross tabulation of agricultural research institutes and responsibility for RDM after leaving the research institute (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
I'll take my data with me and manage it	**A**	8	6.5	10	8.1	8	6.5	4	3.2	5	4.0	**35**	**28.2**
	B	2	1.6	5	4.0	5	4.0	0	0.0	4	3.2	**16**	**12.9**
	C	8	6.5	11	8.9	0	0.0	0	0.0	2	1.6	**21**	**16.9**
	D	2	1.6	3	2.4	5	4.0	3	2.4	1	0.8	**14**	**11.3**
	E	5	4.0	6	4.8	0	0.0	4	3.2	0	0.0	**15**	**12.1**
	F	7	5.6	5	4.0	6	4.8	5	4.0	0	0.0	**23**	**18.5**
My supervisor will be responsible	**A**	10	8.1	13	10.5	11	8.9	0	0.0	1	0.8	**35**	**28.2**
	B	5	4.0	5	4.0	5	4.0	0	0.0	1	0.8	**16**	**12.9**
	C	6	4.8	13	10.5	1	0.8	0	0.0	1	0.8	**21**	**16.9**
	D	1	0.8	5	4.0	6	4.8	0	0.0	2	1.6	**14**	**11.3**
	E	3	2.4	11	8.9	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	10	8.1	8	6.5	5	4.0	0	0.0	0	0.0	**23**	**18.5**
The research institute will assume responsibility	**A**	11	8.9	10	8.1	6	4.8	6	4.8	2	1.6	**35**	**28.2**
	B	7	5.6	2	1.6	2	1.6	5	4.0	0	0.0	**16**	**12.9**
	C	11	8.9	7	5.6	1	0.8	1	0.8	1	0.8	**21**	**16.9**
	D	2	1.6	6	4.8	4	3.2	1	0.8	1	0.8	**14**	**11.3**
	E	5	4.0	9	7.3	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	7	5.6	8	6.5	5	4.0	0	0.0	3	2.4	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes**: A, B, C, D, E, and F

The results in Table 5.17 reveal that 18(14.6%), 7(5.6%), 19(15.4%), 5(4.0%), 11(8.8%) and 12(9.6%) of the respondents from research institutes A, B, C, D, E and F respectively agreed that researchers usually take their data with them after they leave their present institute whereas 9(7.2%) and 5(4.0%) from research institutes A and F in that order disagreed that researchers did not take their data with them if they leave the institute because the institute's policy stipulates that research data belongs to the institute. Another 10(8.0%), 19(15.3%) and 14(11.3%) of the respondents from research institute B, C and E correspondingly stated that the supervisor would be responsible while a significant proposition of respondents 21(17.0%), 18(14.5%) and 14(11.3%) from research institutes respectively A, C and E stated that their research institute will assume responsibility for the data they chose to preserve in its data archive. Less respondents 8(6.5%), 5(4.0%) and 6(4.8%) from research institutes A, D, F in that order were undecided on researchers taking data and manage it themselves after leaving the research institute and 5(4.0%), 1(0.8%) and 1(0.8%) of the respondents from research institutes B. C and E respectively were undecided on their supervisor taking responsibility after leaving the research institute. Further, 6(4.8%) and 5(4.0%) of the respondents form research institutes A and F were undecided on their research institute resuming responsibility after leaving the research institute.

5.5.5 Research data access, sharing and reuse

The study sought to establish how respondents, access, share and reuse research data in the research institute. The results are presented in the following sections 5.5.5.1, 5.5.5.2 and 5.5.5.3.

5.5.5.1 Research data access

Access to research data increases the returns from public investment in research projects; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigator (Organization for Economic Co-operation and Development (OECD), 2007). Effective access to research data in an efficient and effective way is to take advantage of the new opportunities and benefits offered by ICTs. The results are presented in the following sections 5.5.5.1.1, 5.5.5.1.2, 5.5.5.1.3 and 5.5.5.1.4.

5.5.5.1.1 Time taken to halt research data before it is published

The respondents were asked to state the time taken to halt their research data before it is published. The results are shown in Table 5.18.

Table 5.18 Time taken to halt research data before it is published (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
We allow immediate public access to the data	19	15.3	36	29.0	31	25.0	25	20.2	13	10.5	124	100.0
For one year, to permit us to exploit our hard-won research results	22	17.7	45	36.3	31	25.0	17	13.7	9	7.3	124	100.0
Until our journal article has been published	33	26.6	51	41.1	31	25.0	8	6.5	1	0.8	124	100.0

Source: Field data (2017)

Less than half 55(44.3%) of the respondents stated that they allow immediate access to the data whereas 67(54%) asserted that they give a period of one year, to permit them exploit their had-won research results. Further 84(67.7%) of respondents stated that they wait until their journal has been published. Only 9(7.3%) respondents disagreed that they would not halt research data until their journal article is published.

5.5.5.1.2 Ways of restricting public access to research data

The study sought to know restrictions imposed on accessing public research data. The findings are presented in Table 5.19.

Table 5.19 Cross tabulation of agricultural research institutes and ways restricting public access to research data (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Intent to make a patent application thus must avoid prior disclosure	**A**	12	9.7	15	12.1	7	5.6	1	0.8	0	0.0	**35**	**28.2**
	B	6	4.8	6	4.8	4	3.2	0	0.0	0	0.0	**16**	**12.9**
	C	6	4.8	10	8.1	1	0.8	1	0.8	3	2.4	**21**	**16.9**
	D	3	4.8	6	4.8	5	4.0	0	0.0	0	0.0	**14**	**11.3**
	E	5	4.0	10	8.1	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	13	10.5	8	6.5	0	0.0	2	1.6	0	0.0	**23**	**18.5**
The research data are made Confidential by the research group and Commercial partner	**A**	9	7.3	15	12.1	11	8.9	0	0.0	0	0.0	**35**	**28.2**
	B	3	2.4	7	5.6	6	4.8	0	0.0	0	0.0	**16**	**12.9**
	C	6	4.8	14	11.3	0	0.0	0	0.0	1	0.8	**21**	**16.9**
	D	1	0.8	6	4.8	7	5.6	0	0.0	0	0.0	**14**	**11.3**
	E	2	1.6	12	9.7	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	13	10.5	8	6.5	0	0.0	2	1.6	0	0.0	**23**	**18.5**
Confidential, proprietary or classified research data	**A**	6	4.8	15	12.1	12	9.7	2	1.6	0	0.0	**35**	**28.2**
	B	3	2.4	7	5.6	5	4.0	1	0.8	0	0.0	**16**	**12.9**
	C	5	4.0	13	10.5	2	1.6	0	0.0	1	0.8	**21**	**16.9**
	D	1	0.8	3	2.4	10	8.1	0	0.0	0	0.0	**14**	**11.3**
	E	2	1.6	13	10.5	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	11	8.9	10	8.1	0	0.0	2	1.6	0	0.0	**23**	**18.5**
Intellectual property concerns	**A**	6	4.8	13	10.5	10	8.1	5	4.0	1	0.8	**35**	**28.2**
	B	4	3.2	5	4.0	6	4.8	1	0.8	0	0.0	**16**	**12.9**
	C	7	5.6	9	7.3	1	0.8	3	2.4	1	0.8	**21**	**16.9**
	D	2	1.6	3	2.4	6	4.8	1	0.8	2	1.6	**14**	**11.3**
	E	5	4.0	9	7.3	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	8	6.5	13	10.5	0	0.0	2	1.6	0	0.0	**23**	**18.5**
Lack of appropriate tools for accessing or publishing data	**A**	7	5.6	10	8.1	10	8.1	6	4.8	2	1.6	**35**	**28.2**
	B	3	2.4	5	4.0	6	4.8	1	0.8	1	0.8	**16**	**12.9**
	C	5	4.0	11	8.9	1	0.8	2	1.6	2	1.6	**21**	**16.9**

D	2	1.6	1	0.8	8	6.5	2	1.6	1	0.8	**14**	**11.3**
E	0	0.0	10	8.1	0	0.0	3	2.4	2	1.6	**15**	**12.1**
F	13	10.5	5	4.0	0	0.0	5	4.0	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural Research institutes:** A, B, C, D, E, and F

The results reveal that majority 27(21.8%), 16(12.9%), 15(12.1%) and 21(17.0%) of the respondents from research institutes A, C, E and F respectively stated that they intend to make a patent application and subsequently must avoid prior disclosure. The research data are confidential because of the arrangements the research groups have made with the commercial partner sponsoring their research. This was according to 10(8.0%), 20(16.1%) and 14(11.4) of the respondents from research institutes B, C and E in that order. According to 21(16.9%) 10(8.0%) and 18(14.5%) of the respondents from research institutes A, B and C correspondingly, the research data can be protected from public access by being categorized as confidential, proprietary or classified information. Another 5(4.0%), 14(11.3% and 21(17.0%) of the respondents from research institutes D, E and F noted that research data can be restricted from public access through intellectual property concerns while 1(0.8%) and 2(1.6%) of the respondents from research institutes A and D respectively disagree that they cannot restrict public access through intellectual property whereas 6(4.8% and 1(0.8%) of respondents from research institutes D and E respectively were undecided on restricting research data through intellectual property. The research data access may also be limited through lack of appropriate tools for sharing or publishing data as stated by majority of the respondents 17(13.7%), 15(12.0%) and 18(14.5%) from research institutes A, C and F in that order whereas 8(6.4%), 4(3.2%) and 5(4.0%) of respondents from the same research institutes disagreed that research data cannot be limited through lack of appropriate tools for sharing or publishing data.

5.5.5.1.3 Accessing research data

The study also sought to determine the persons who were allowed to access the research data. Figure 5.5 shows the results.

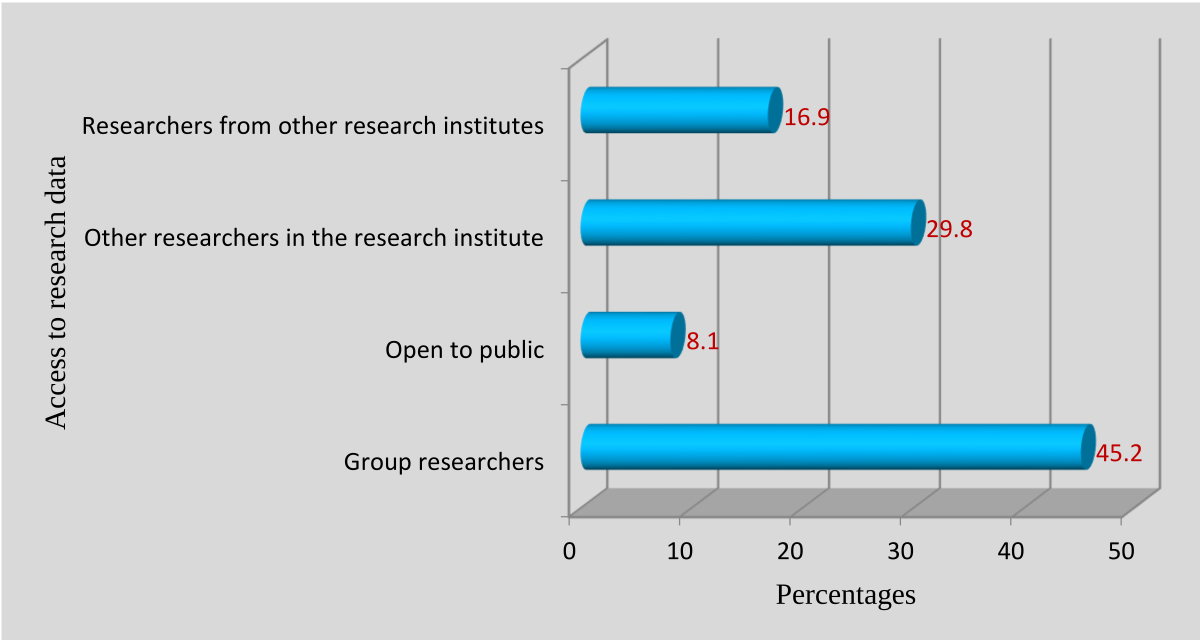

Figure 5.5 Access to research data (n=124)

As indicated by the results presented in Fig. 5.5, 56(45.2%) respondents indicated that group researchers were allowed to access research data. However, 37(29.8%) and 21(16.9%) respondents revealed that other researchers in the research institute and researchers from other research institute respectively were currently accessing the research data. The research data was also open to public as stated by 10(8.1%) respondents.

5.5.5.1.4 Availing research data through open access

The respondents were also asked to state how they availed research data through open access. The results are shown in Figure 5.6.

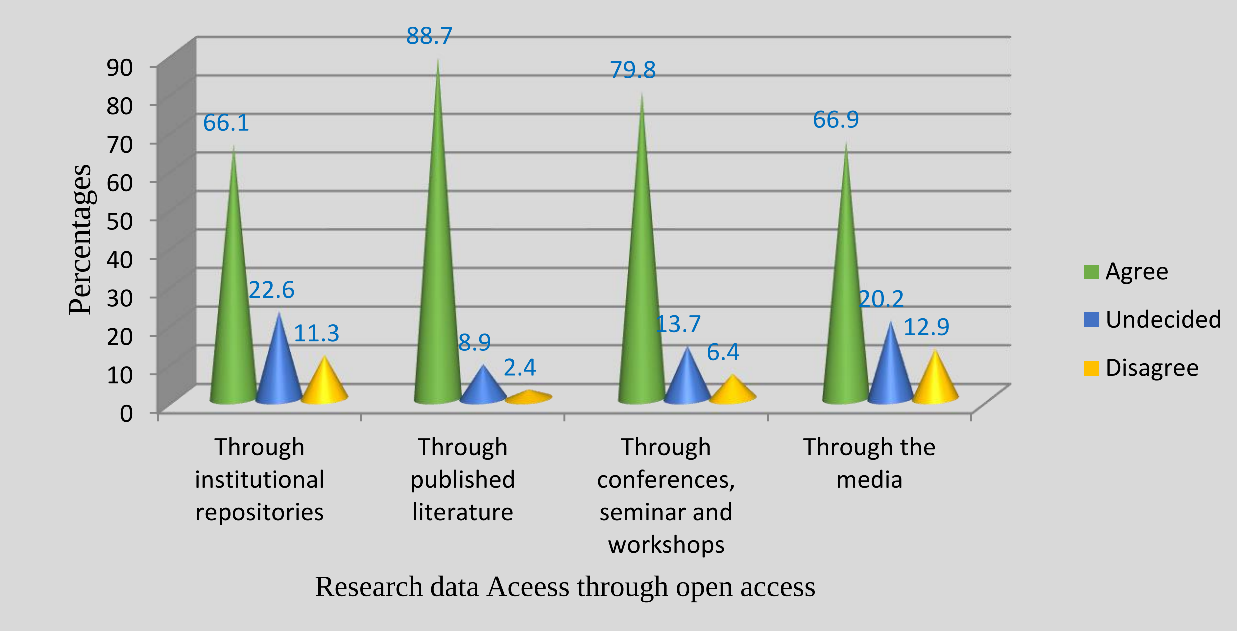

Figure 5.6 Research data access through open access (n=124)

The results presented in Fig.5.6 revealed that 82(66.1%) of the respondents avail research data through institutional repositories whereas 110(88.7%) availed the data through publishing literature and 99(79.8%) availed the research data through conference, seminars and workshops. Only 83(66.9%) availed their research data through the media. There were 16(12.9%) respondents who disagreed that they did not avail research data through the media

5.5.5.2 Sharing research data

Data sharing has incredible potential to strengthen agricultural research. Warren (2016) acknowledges that the benefits of data sharing is attained when researchers have access to complete datasets and are thus able to answer new questions, explore different lines of analysis, and more efficiently conduct large-scale analyses in a research. This section will address questions on data sharing

5.5.5.2.1 Sharing research data before publication of any papers

The respondents were asked to state the persons they shared the research data with before publication of any papers. The results are shown in Table 5.20.

Table 5.20 Cross tabulation of agricultural research institutes and whom they share research data with (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
My research Supervisor	**A**	10	8.1	9	7.3	10	8.1	6	4.8	0	0.0	**35**	**28.2**
	B	7	5.6	1	0.8	7	5.6	1	0.8	0	0.0	**16**	**12.9**
	C	4	3.2	14	11.3	2	1.6	0	0.0	1	0.8	**21**	**16.9**
	D	2	1.6	6	4.8	3	2.4	2	1.6	1	0.8	**14**	**11.3**
	E	3	2.4	9	7.3	1	0.8	2	1.6	0	0.0	**15**	**12.1**
	F	8	6.5	11	8.9	4	3.2	0	0.0	0	0.0	**23**	**18.5**
Members of my research group and trusted external Collaborators	**A**	9	7.3	17	13.7	4	3.2	3	2.4	2	1.6	**35**	**28.2**
	B	6	4.8	7	5.6	0	0.0	2	1.6	1	0.8	**16**	**12.9**
	C	7	5.6	10	8.1	1	0.8	0	0.0	3	2.4	**21**	**16.9**
	D	4	3.2	4	3.2	4	3.2	1	0.8	1	0.8	**14**	**11.3**
	E	5	4.0	9	7.3	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	8	6.5	11	8.9	4	3.2	0	0.0	0	0.0	**23**	**18.5**
Research sponsors	**A**	12	9.7	9	7.3	4	3.2	4	3.2	6	4.8	**35**	**28.2**
	B	9	7.3	3	2.4	0	0.0	0	0.0	4	3.2	**16**	**12.9**
	C	5	4.0	11	8.9	1	0.8	1	0.8	3	2.4	**21**	**16.9**
	D	2	1.6	3	2.4	3	2.4	4	3.2	2	1.6	**14**	**11.3**
	E	4	3.2	9	7.3	2	1.6	0	0.0	0	0.0	**15**	**12.1**
	F	8	6.5	11	8.9	2	1.6	2	1.6	0	0.0	**23**	**18.5**
General public	**A**	4	3.2	13	10.5	6	4.8	3	2.4	9	7.3	**35**	**28.2**
	B	3	2.4	7	5.6	2	1.6	0	0.0	4	3.2	**16**	**12.9**
	C	2	1.6	10	8.1	5	4.0	2	1.6	2	1.6	**21**	**16.9**
	D	2	1.6	2	1.6	4	3.2	3	2.4	3	2.4	**14**	**11.3**
	E	2	1.6	7	5.6	1	0.8	2	1.6	3	2.4	**15**	**12.1**
	F	2	1.6	14	11.3	4	3.2	0	0.0	3	2.4	**23**	**18.5**
Everyone, by publishing the data Online	**A**	4	3.2	13	10.5	3	2.4	5	4.0	10	8.1	**35**	**28.2**
	B	1	0.8	7	5.6	1	0.8	0	0.0	7	5.6	**16**	**12.9**
	C	3	2.4	10	8.1	4	3.2	1	0.8	3	2.4	**21**	**16.9**
	D	2	1.6	4	3.2	2	1.6	4	3.2	2	1.6	**14**	**11.3**
	E	5	4.0	8	6.5	1	0.8	1	0.8	0	0.0	**15**	**12.1**
	F	2	1.6	7	5.6	5	4.0	6	4.8	3	2.4	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes**: A, B, C, D, E, and F

The results show that the respondents (researchers) shared research data with their research supervisors 19(15.4%), 8(6.45) and 18(14.5%) from research institutes A, B and C respectively agreed. On the contrary, 6(4.8%), 1(0.8%), 1(0.8%), 3(2.4%) and 2(1.6%) of respondents from research institutes A, B, C, D and E in that order disagreed that they did not share research with their research supervisor. However, a significant number of respondents 26(21.0%), 17(15.4%) and 19(13.7%) from research institutes A, C and F correspondingly agreed that they shared research data with members of their research groups and trusted external collaborators while 4(3.2%) of the respondents each from research institutes A, D and F were undecided on sharing research data with members of their research groups and trusted external collaborators. There were 5(4.0%) and 13(10.5%) respondents from research institutes D and E in that order who agreed that they shared research data with research sponsors, 17(13.7%) and 16(12.9%) from research institutes A and F respectively shared research data with the general the public and 8(6.4%) and 13(10.5%) from research institutes B and C correspondingly shared research data with everyone by publishing research data online. Only 15(12.1%), 7(5.6%), 4(3.2%), 6(4.8%), 1(0.8%) and 9(7.2%) of the respondents from research institutes A, B, C, D, E and F correspondingly disagreed that they did not share research data with everyone by publishing research data online. In relation to whom they shared research data, one respondent [HR13] interviewed, pointed out that 'raw' data was rarely shared but could be shared through publications. However, it all depended on the level of research data to be shared. The respondent further lamented that it was risky to share because other researchers could plagiarize such data.

5.5.5.2.2 Method of sharing all or part of research data

The study wanted to investigate the method of sharing all or part of research data. Table 5.21 shows the methods used to share all or part of the research data.

Table 5.21 Methods of sharing all or part of research data (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Collaborative web space (wiki, blog, Google Docs)	25	20.2	49	39.5	23	18.5	14	11.3	13	10.5	124	100.0
Data portal or database driven web site	29	23.4	47	37.9	33	26.6	8	6.5	7	5.6	124	100.0
Deposit them with a specialist data centre	24	19.4	50	40.3	27	21.8	15	12.1	8	6.5	124	100.0
Depositing them in an institutional repository	34	27.4	43	34.7	31	25.0	9	7.3	7	5.6	124	100.0
Submitting them to a journal to support publication	35	28.2	42	33.9	40	32.3	1	0.8	6	4.8	124	100.0
E-mail	37	29.8	55	44.4	20	16.1	6	4.8	6	4.8	124	100.0

External storage device (USB drive, CD/DVD)	50	40.3	50	40.3	13	10.5	3	2.4	8	6.5	124	100.0
Hard copy or print	50	40.3	44	35.5	12	9.7	9	7.3	9	7.3	124	100.0
Don't share data	46	37.1	43	34.7	15	12.1	9	7.3	11	8.9	124	100.0

Source: Field data (2017)

The results show that 74(59.7%) of the respondents agreed that they shared their research data through collaborative web space, 76(61.3%) shared research data through data portal or database drive website, 74(59.7%) shared research data by depositing it with a specialized data centre, 77(62.1%) shared the research data by depositing them in an institutional repository, 77(62.1%) respondents submitted their research data to a journal for publication and 92(74.2%) shared their raw research data through email. The findings also indicate that 100(80.6%) of the respondents shared their research data through an external storage device while 94(75.8%) shared their research data through hard copy or print. Furthermore 89(71.8%) of the respondents stated that they didn't share data and 20(16.2%) disagreed that they did share research data.

5.5.5.2.3 Benefits of sharing research data

The study sought to establish the benefits of sharing research data. The responses are presented in Table 5.22.

Table 5.22 Benefits of sharing research data (n=124)

Statement	**SA**		**A**		**U**		**D**		**SD**		**Total**	
	f	**%**	**f**	**%**	**f**	**%**	**f**	**%**	**F**	**%**	**f**	**%**
Promote innovations and potential new data uses	44	35.5	44	35.5	17	13.7	2	1.6	17	13.7	124	100.0
Encourages scientific enquiry and debate	61	49.2	49	39.5	8	6.5	1	0.8	5	4.0	124	100.0
Reduce the cost of duplicating data collection	59	47.6	51	41.1	6	4.8	4	3.2	4	3.2	124	100.0
Enables scrutiny of research findings	47	37.9	59	47.6	13	10.5	2	1.6	3	2.4	124	100.0
Increases the impact and visibility of research	43	34.7	65	52.4	7	5.6	6	4.8	3	2.4	124	100.0
Leads to new collaborations	52	41.9	52	41.9	14	11.3	3	2.4	3	2.4	124	100.0

Source: Field data 2017

The results in Table 5.22 show that 88(71%) of the respondents stated that sharing research data promotes innovations and potential new data uses while 110(88.7%) stated that sharing of research data encourages scientific enquiry and debate. Furthermore, 110(88.7%) respondents agreed that sharing research data reduces the cost of duplicating data collection whereas 106(85.5%) stated that data sharing enables scrutiny of research findings. Data sharing also increases the impact and visibility of research as stated by 108(87.1%) respondents whereas 9(7.2%) of the respondents disagreed that research data did not increase the impact and visibility of research. Respondents who stated that sharing research data leads to new collaborations between data users and data creators were 104(83.8%), 14(11.3%) were undecided on research data leading to new collaborations and 6(4,8%) of the respondents disagreed that research data did not lead to new collaborations. These findings concur with the interview results where the heads of research who were interviewed stated that sharing research data brings accessibility to information seekers and creation of awareness. It also enhances the dissemination of information to clients/stakeholders within and outside the country.

5.5.5.2.4 Challenges in research data sharing

The study sought to establish the Challenges encountered during research data sharing. The responses are presented in Figure 5.7.

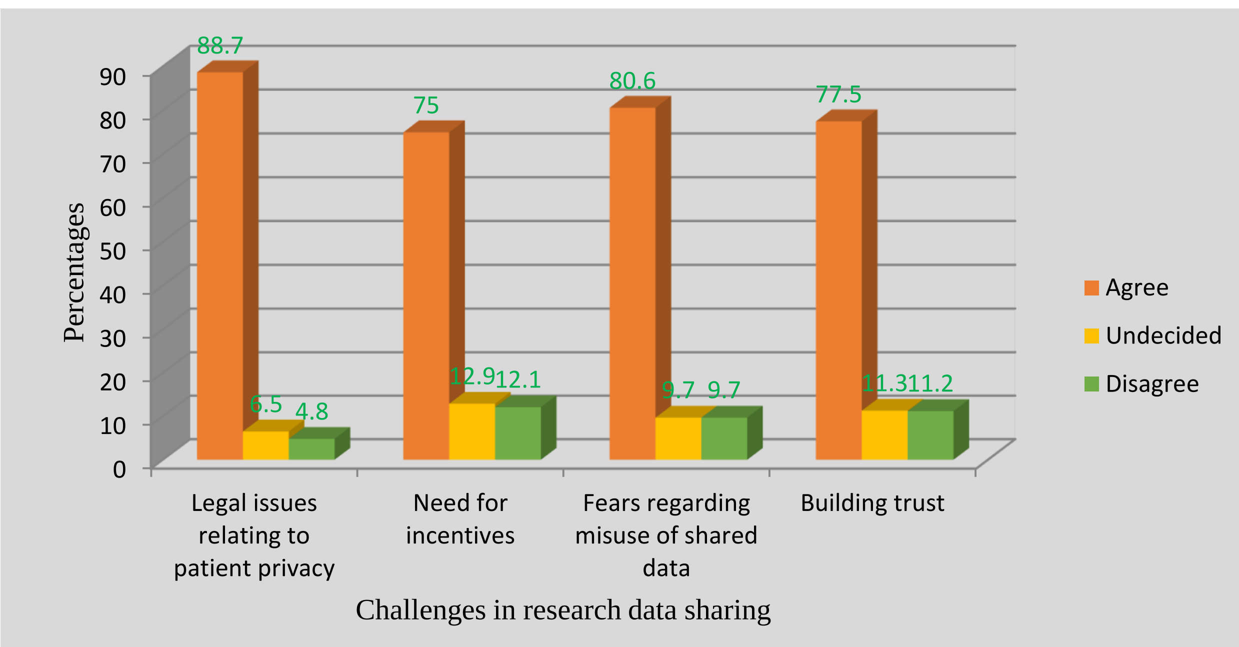

Figure 5.7 Challenges in research data sharing (n=124)

The results shown in Figure 5.7 reveal that, 110(88.7%) stated that legal issues relating to patent privacy was a major challenge in data sharing. Further 93(75%) of the respondents stated that there is need for incentives in order to share research data. Similarly 100(80.6%) of the respondents stated that there were fears regarding misuse of shared data and 96(77.4%) stated that there was a challenge of building trust. Head of research [HR 15] from Research institute E interviewed, noted that the challenges scientists face (researchers) are fear of losing data to plagiarism, lack of trust among researchers, lack of facilitations to share data for example ICT tools and equipment, absence of data literacy and advocacy programs, to mention a few , hamper sharing of research data.

5.5.5.3 Reusing research data

Research data can be reused in increasingly diverse ways and have potential beyond the original scope of a research project. This includes the data that may have been collected for a previous study that can be used to support new research.

5.5.5.3.1 Using and reusing research data

The respondents were asked to state the reasons for using and reusing research data. The results are shown in Table 5.23.

Table 5.23 Reasons for using and reusing data (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Avoid duplication	52	41.9	49	39.5	11	8.9	2	1.6	10	8.1	124	100.0
Reduce the cost of doing research	55	44.4	43	34.7	16	12.9	1	0.8	9	7.3	124	100.0
Re-analysis of data can lead to powerful insights	57	46.0	56	45.2	3	2.4	5	4.0	3	2.4	124	100.0
Encourages scientific enquiry	63	50.8	51	41.1	4	3.2	2	1.6	4	3.2	124	100.0
Promote innovations	52	41.9	60	48.4	6	4.8	5	4.0	1	0.8	124	100.0
Potential new data uses	59	47.6	57	46.0	4	3.2	3	2.4	1	0.8	124	100.0

Source: Field data (2017)

The results presented in Table 5.23 revealed that research data is used and reused in order to avoid duplication as stated by 101(81.4%) respondents whereas 12(9.7%) disagreed that using and re-using did not avoid duplication. Additionally, 11(8.9%) of the respondents were undecided on using and re-using research data in avoiding duplication. Furthermore, 98(79.1%) agreed that using and re-using research data reduces the cost of doing research while 113(91.2%) agreed that re-analysis of research data can lead to powerful insights and 8(6.4%) disagreed that re-analysis of research data did not lead to powerful insights. Additionally, 114(91.1%), 112(90.3%) and 116(93.6%) respondents agreed that using and re-using research data encourages scientific inquiry, promotes innovations and encourages potential new data uses respectively. Consequently, the using and re-using of research data can be summarized in the testimony of one head of research [HR 9] who had this to say:

> *"Data is always reused especially in agricultural research institutes. Sometimes when doing a report on certain research project, one may reuse the data if results are not tallying as expected. A case in point is the scenario where data sheets done by National Gene bank of Kenya in 1988 are still used and reused till today for reference. Also reuse of coffee research data as gone a long way in improving variety of coffee and even bringing in new varieties that are manageable to the farmers like the new variety called ruiru coffee. Food crops have equally improved their varieties due to reuse of data. Re-using of data allow analysis of new areas of research.*

Based on the results, it is clear that re-using of research data plays a key role in advancing research and discovering new insights.

5.5.5.3.2 Ways of making research data available for use and reuse

Findings on the ways of making research data available for use and reuse are presented in Figure 5.24.

Table 5.24 Cross tabulation of agricultural research institutes and ways of making research data available for use and reuse (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Publication	**A**	15	12.1	17	13.7	2	1.6	1	0.8	0	0.0	**35**	**28.2**
	B	6	4.8	9	7.3	1	0.8	0	0.0	0	0.0	**16**	**12.9**
	C	8	6.5	11	8.9	0	0.0	1	0.8	1	0.8	**21**	**16.9**
	D	4	3.2	6	4.8	4	3.2	0	0.0	0	0.0	**14**	**11.3**
	E	11	8.9	4	3.2	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	11	8.9	7	5.6	0	0.0	3	2.4	2	1.6	**23**	**18.5**
Citation	**A**	18	14.5	11	8.9	4	3.2	1	0.8	1	0.8	**35**	**28.2**
	B	10	8.1	5	4.0	1	0.8	0	0.0	0	0.0	**16**	**12.9**
	C	12	9.7	6	4.8	1	0.8	2	1.6	0	0.0	**21**	**16.9**
	D	6	4.8	4	3.2	4	3.2	0	0.0	0	0.0	**14**	**11.3**
	E	7	5.6	7	5.6	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	18	14.5	0	0.0	0	0.0	5	4.0	0	0.0	**23**	**18.5**
Sufficient metadata	**A**	11	8.9	18	14.5	2	1.6	3	2.4	1	0.8	**35**	**28.2**
	B	9	7.3	7	5.6	0	0.0	0	0.8	0	0.0	**16**	**12.9**
	C	4	3.2	15	12.1	1	0.8	1	0.8	0	0.0	**21**	**16.9**
	D	4	3.2	7	5.6	2	1.6	1	0.8	0	0.0	**14**	**11.3**
	E	3	2.4	9	7.3	1	0.8	1	0.8	1	0.8	**15**	**12.1**
	F	5	4.0	13	10.5	0	0.0	3	2.4	2	1.6	**23**	**18.5**
Research data licensing approaches	**A**	13	10.5	10	8.1	8	6.5	3	2.4	1	0.8	**35**	**28.2**
	B	10	8.1	3	2.4	3	2.4	0	0.0	0	0.0	**16**	**12.9**
	C	3	2.4	15	12.1	2	1.6	0	0.0	1	0.8	**21**	**16.9**
	D	3	2.4	5	4.0	4	3.2	0	0.0	2	1.6	**14**	**11.3**
	E	2	1.6	8	6.5	3	2.4	2	1.6	0	0.0	**15**	**12.1**
	F	4	3.2	8	6.5	3	2.4	0	0.0	8	6.5	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes:** A, B, C, D, E, and F.

The results reveal that the majority of the respondents agreed that research data should be made available for use and reuse through publications. This was according to 32(25.8%) and 19(15.4%)

of the respondents from research institutes A and C correspondingly. Respondents 15(12.1%) and 14(11.2%) from research institutes B and E in that order indicated that research data is made available through citation, and 11(8.8%) and 18(14.5%) respondents from research institutes D and F in that order pointed out that research data is made available through sufficient metadata describing how the data has been specified, collected, analyzed and transformed. However 23(18.6%), 13(10.5%), 18(14.5%), 8(6.4%), 10(8.1%) and 12(9.7%) of the respondents from research institutes A, B, C, D, E and F correspondingly agreed that research data was made available through research data licensing approaches while 2(1.6%) of the respondents each from research institutes D and E disagreed that research data was not made available through research data licensing approaches for use and reuse.

5.5.5.4 Responsibility for RDM in research institute

The study also sought to find out the person who was responsible for the research data management in the research institute. The findings are shown in Table 5.25.

Table 5.25 Cross tabulation of agricultural research institutes and responsibility for RDM in research institute (n=124)

Statement		SA		A		U		D		SD		Total	
		f	%	f	%	f	%	f	%	f	%	F	%
IT staff within research institute	**A**	3	2.4	10	8.1	5	4.0	10	8.1	7	5.6	**35**	**28.2**
	B	3	2.4.	6	4.8	2	1.6	4	3.2	1	0.8	**16**	**12.9**
	C	3	2.4	11	8.9	4	3.2	1	0.8	2	1.6	**21**	**16.9**
	D	1	0.8	4	3.2	1	0.8	5	4.0	3	2.4	**14**	**11.3**
	E	0	0.0	7	5.6	4	3.2	0	0.0	4	3.2	**15**	**12.1**
	F	2	1.6	8	6.5	6	4.8	3	2.4	4	3.2	**23**	**18.5**
Librarian	**A**	3	2.4	10	8.1	12	9.7	3	2.4	7	5.6	**35**	**28.2**
	B	2	1.6	4	3.2	9	7.3	0	0.0	1	0.8	**16**	**12.9**
	C	1	0.8	9	7.3	6	4.8	2	1.6	3	2.4	**21**	**16.9**
	D	1	0.8	3	2.4	3	2.4	3	2.4	4	3.2	**14**	**11.3**
	E	2	1.6	4	3.2	4	3.2	0	0.0	5	4.0	**15**	**12.1**
	F	2	1.6	3	2.4	0	0.0	8	6.5	10	8.1	**23**	**18.5**
Collaborative responsibility or research group	**A**	2	1.6	12	9.7	14	11.3	5	4.0	2	1.6	**35**	**28.2**
	B	2	1.6	3	2.4	10	8.1	1	0.8	0	0.0	**16**	**12.9**
	C	2	1.6	13	10.5	5	4.0	0	0.0	1	0.8	**21**	**16.9**
	D	1	0.8	3	2.4	3	2.4	5	4.0	2	1.6	**14**	**11.3**
	E	1	0.8	12	9.7	1	0.8	0	0.0	1	0.8	**15**	**12.1**
	F	4	3.2	9	7.3	0	0.0	5	4.0	5	4.0	**23**	**18.5**
External research partners	**A**	5	4.0	10	8.1	12	9.7	6	4.8	2	1.6	**35**	**28.2**
	B	3	2.4	3	2.4	9	7.3	1	0.8	0	0.0	**16**	**12.9**

	C	4	3.2	11	8.9	1	0.8	1	0.8	4	3.2	**21**	**16.9**
	D	1	0.8	2	1.6	1	0.8	7	5.6	3	2.4	**14**	**11.3**
	E	1	0.8	9	7.3	1	0.8	0	0.0	4	3.2	**15**	**12.1**
	F	2	1.6	17	13.7	0	0.0	2	1.6	2	1.6	**23**	**18.5**
Third-party data center	**A**	1	0.8	8	6.5	13	10.5	6	4.8	7	5.6	**35**	**28.2**
	B	0	0.0	3	2.4	11	8.9	1	0.8	1	0.8	**16**	**12.9**
	C	0	0.0	13	10.5	0	0.0	1	0.8	7	5.6	**21**	**16.9**
	D	1	0.8	2	1.6	1	0.8	6	4.8	4	3.2	**14**	**11.3**
	E	0	0.0	7	5.6	0	0.0	0	0.0	8	6.5	**15**	**12.1**
	F	2	1.6	6	4.8	0	0.0	11	8.9	4	3.2	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes**: A, B, C, D, E, and F

The results in Table 5.24 show that 13(10.5%) and 14(11.3%) of respondents from research institutes A and C respectively stated that IT staff within research institute were in-charge of RDM whereas 6(4.8%) of respondents each from research institutes B and E correspondingly stated that librarians were responsible. Additionally 4(3.2%) and 13(10.5%) of the respondents from research institutes D and F in that order stated that collaborative responsibility or research groups were responsible for RDM whereas 6(4.8%) and 15(12.1%) respondents from research institutes B and C respectively indicated that external research partners were responsible for RDM. Another 7(5.6%) and 8(6.4%) of respondents from research institutes E and F stated that third-party data centre were in-charge of RDM whereas 8(6.4%) and 15(12.1%) of the respondents from the same research institutes disagreed that the third-party data centre was not responsible for RDM in research institutes.

During the interview, the heads of research, IT and librarians were asked to state how research data was captured**,** appraised, described, preserved, accessed and reused in their institute. The findings from interviews revealed that data capture was done differently in different institutes. There were institutes where data capture was by and large paper work, where the questionnaires are prepared and filled manually by the respondents. In other institutes, there were field attendants who were used to capture notes through daily observation. They incorporate a scheduled period for data collection, report compilation from data forms, and then save it as soft copies on computers and backup devices and also as hard copies. Excel sheets are shared but the manual forms are stored for reference, and then published. One of the heads of research [HR4] commented:

> "*…we first do a pretest and if it accepted according to SOPs, then data is capture using appropriate means. Finally data is capture in soft copy and printed later to*

be preserved in the files and cabinets. Data collected is then cleaned up to get useful data with all the files labeled. To preserve research data, researchers are supposed to deposit in the institutional repository…..which rarely happens"

"…..The responsibility of RDM currently in the institute is head of research but RDM units should be established to coordinate from all the departments to allow efficient access and sharing of data"

The respondents who were interviewed also stated that data was always reused for benchmarking on what was done before and what is being done now and there was a need for all data/files to be titled for easy identification. In some institutes where the study was done, there was a standard format for publication, specification and guidelines, in the form of a thesis. The findings from the interview also revealed that the prepared research report was shared among the committee members and each committee member gives feedback and the report is then reviewed where the committee decides which data to keep. The committee then decides if the data is relevant or not.

5.6 RDM knowledge, skills and training requirements

Community Capability Model (CCM) framework (Lyon et al., 2012) is used to underpin the study suitable to its capabilities in RDM, likewise to the research question. RDM involves services, tools and infrastructure that support the management of research data across all stage in the lifecycle. Essentially, data curation and RDM capabilities need to be developed for an effective RDM in agricultural research institute (Creamer, Morales and Crespo, 2012). The DCC lifecycle model (Higgins, 2008) which also informs the study, ensures that it breaks down the range of roles, responsibilities and actions in every stage of data curation lifecycle that need to be considered when developing competencies in terms of knowledge, skills, and training required by RDM stakeholders. The model resonates with CCM model framework on knowledge, skill and training required for RDM. The same DCC lifecycle model has also been used to inform the development of digital curation curricula in institutions of higher learning to enhance the competencies required in RDM. The Cronbach's Alpha values for this question were 0.698 for items in the researcher's questionnaire. This suggested a high internal validity of the test items. The research question therefore sought to find out RDM knowledge, skills and training needed to capture, appraise, describe, preserve, access and reuse its research data. Questions D (1-3) of the questionnaire for researchers (Appendix 4), questions D (1-6) of the interview schedule for directors of institute

(Appendix 1), questions D (1-5) of the interview schedule for heads of research (Appendix 2), and questions D (1-5) of the interview schedule for heads of IT and librarians (Appendix 3) addressed this research question. The results are presented in the following sections 5.6.1, 5.6.2, 5.6.3, 5.6.4 and 5.6.5

5.6.1 Knowledge needs for RDM

The researchers were first asked to state the type of knowledge they posses for research data capture, appraisal, preservation, access and reuse in the research institute. The responses are shown in Table 5.26.

Table 5.26 Type of RDM knowledge possessed (n=124)

Statement	HS		S		N		LS		NS		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Data Curation (capture, appraisal, description)	22	17.7	60	48.4	24	19.4	12	9.7	6	4.8	124	100.0
Preservation	20	16.1	66	53.2	26	21.0	6	4.8	6	4.8	124	100.0
Open access	25	20.2	69	55.6	16	12.9	9	7.3	5	4.0	124	100.0
Sharing	23	18.5	68	54.8	18	14.5	10	8.1	5	4.0	124	100.0

Source: Field data (2017)

The results presented in Table 5.25 show that, 82(66.1%) of the respondents stated that they possessed knowledge in data curation (capture, appraisal and description) whereas 86(69.3%) stated that they have knowledge in data preservations. Further, 26 (21%) of the respondents were neutral on possessing preservation skill, whereas 6(4.8%) of the respondents had less skill in preservation and another 6(4.8%) had no skills on the same. The findings also show that 94(75.8%) of the respondents stated that they have knowledge in open access while 91(73.3%) had knowledge in data sharing. However, 9(7.3%) of the respondents stated that they were less skilled in open access and 5(4.0%) indicated that they had no skills in sharing research data.

5.6.2 RDM skills possessed

There was also a need to determine the skills that the respondents possessed for research data capture, appraisal, description, preservation, access and reuse in their research institute. The responses are shown in Table 5.27.

Table 5.27 RDM skills possessed (n=124)

Statement		HS		S		N		LS		NS		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
Metadata skills	**A**	5	4.0	12	9.7	5	4.0	11	8.9	2	1.6	**35**	**28.2**
	B	0	0.0	7	5.6	3	2.4	5	4.0	1	0.8	**16**	**12.9**
	C	5	4.0	12	9.7	0	0.0	2	1.6	2	1.6	**21**	**16.9**
	D	2	1.6	5	4.0	1	0.8	6	4.8	0	0.0	**14**	**11.3**
	E	5	4.0	4	3.2	1	0.8	2	1.6	3	2.4	**15**	**12.1**
	F	3	2.4	13	10.5	2	1.6	0	0.0	5	4.0	**23**	**18.5**
Searching and retrieval skills	**A**	6	4.8	13	10.5	8	6.5	8	6.5	0	0.0	**35**	**28.2**
	B	2	1.6	5	4.0	5	4.0	4	3.2	0	0.0	**16**	**12.9**
	C	6	4.8	11	8.9	2	1.6	1	0.8	1	0.8	**21**	**16.9**
	D	2	1.6	5	4.0	1	0.8	6	4.8	0	0.0	**14**	**11.3**
	E	3	2.4	12	9.7	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	6	4.8	9	7.3	2	1.6	0	0.0	6	4.8	**23**	**18.5**
Data capturing skills	**A**	8	6.5	12	9.7	9	7.3	5	4.0	1	0.8	**35**	**28.2**
	B	2	1.6	5	4.0	6	4.8	3	2.4	0	0.0	**16**	**12.9**
	C	3	2.4	12	9.7	4	3.2	1	0.8	1	0.8	**21**	**16.9**
	D	3	2.4	6	4.8	2	1.6	3	2.4	0	0.0	**14**	**11.3**
	E	7	5.6	8	6.5	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	8	6.5	10	8.1	2	1.6	0	0.0	3	2.4	**23**	**18.5**
Preservation skills	**A**	6	4.8	13	10.5	10	8.1	6	4.8	0	0.0	**35**	**28.2**
	B	1	0.8	6	4.8	5	4.0	4	3.2	0	0.0	**16**	**12.9**
	C	5	4.0	14	11.3	1	0.8	0	0.0	1	0.8	**21**	**16.9**
	D	3	2.4	2	1.6	4	3.2	5	4.0	0	0.0	**14**	**11.3**
	E	7	5.6	8	6.5	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	13	10.5	7	5.6	0	0.0	0	0.0	3	2.4	**23**	**18.5**
Sharing skill	**A**	5	4.0	15	12.1	7	5.6	8	6.5	0	0.0	**35**	**28.2**
	B	4	3.2	5	4.0	4	3.2	4	3.2	0	0.0	**16**	**12.9**
	C	6	4.8	12	9.7	3	2.4	0	0.0	0	0.0	**21**	**16.9**
	D	1	0.8	7	5.6	1	0.8	5	4.0	0	0.0	**14**	**11.3**
	E	3	2.4	12	9.7	0	0.0	0	0.0	0	0.0	**15**	**12.1**
	F	11	8.9	10	8.1	2	1.6	0	0.0	0	0.0	**23**	**18.5**
Tools and Technologies	**A**	4	3.2	17	13.7	8	6.5	5	4.0	1	0.8	**35**	**28.2**
	B	1	0.8	7	5.6	4	3.2	4	3.2	0	0.0	**16**	**12.9**
	C	3	2.4	14	11.3	4	3.2	0	0.0	0	0.0	**21**	**16.9**
	D	5	4.0	3	2.4	3	2.4	2	1.6	1	0.8	**14**	**11.3**
	E	3	2.4	11	8.9	1	0.8	0	0.0	0	0.0	**15**	**12.1**
	F	11	8.9	10	8.1	0	0.0	2	1.6	0	0.0	**23**	**18.5**
Collaboration and communication	**A**	4	3.2	19	15.3	6	4.8	4	3.2	1	0.8	**35**	**28.2**
	B	1	0.8	7	5.6	4	3.2	4	3.2	0	0.0	**16**	**12.9**
	C	6	4.8	14	11.3	1	0.8	0	0.0	0	0.0	**21**	**16.9**
	D	3	2.4	7	5.6	2	1.6	1	0.8	1	0.8	**14**	**11.3**
	E	1	0.8	10	8.1	0	0.0	3	2.4	1	0.8	**15**	**12.1**

F	11	8.9	8	6.5	0	0.0	4	3.2	0	0.0	**23**	**18.5**

Source: Field data (2017); **Agricultural research institutes**: A, B, C, D, E, and F

The results shown in Table 5.26 reveal that 17(13.7%) and 7(5.6%) of the respondents from research institutes A and D respectively stated that they possessed metadata skills whereas 11(8.9%) and 6(4.8%) indicated that they had less skills, 2(1.6%) and 0(0.0%) of the respondents had no metadata skills and 5(4.0%) and 1(0.8%) were undecided on possessing metadata skill all from the same group of respondents. There were 7(5.6%) and 15(12.1%) of the respondents from research institutes B and F who stated that they had searching and retrieval skills while 15(12.1%) of the respondents, each from research institutes C and E had data capturing skills. From the same group of respondents 1(0.8%) and 0(0.0%) indicated they had less skills in capturing skills. Furthermore, 19(15.3%), 20(16.1%) and 21(16.9%) of the respondents from research institute A had data preservation skills, sharing skills and tools and technologies respectively. Another 10(8.0%), 1(0.8%), and 1(0.8%) respondents from research institute D had skills, less skills and no skills respectively in collaboration and communication

During the interview with the heads of research, librarian, and IT specialists, it was established that the researchers and the heads of departments needed training in library and IT skills such as metadata, searching and retrieval, and new tools and technologies which would enable them to be efficient and effective in RDM. For one to be able to do research, data capture, appraisal, description, preservation, access and reuse, they needed training in management and dissemination of data, knowledge of RDM tools and web management.

5.6.2.1 Skills gap in RDM

Another concern for this study was to determine the areas in research data management where there were skills gaps. The results are presented in Table 5.28.

Table 5.28 Skills gap in RDM (n=124)

Statement	HS		S		N		LS		NS		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Developing metadata schema	55	44.4	44	35.5	14	11.3	8	6.5	3	2.4	124	100.0
Use of Information and Communication Technologies (ICTs) tools and equipment’s	47	37.9	54	43.5	12	9.7	9	7.3	2	1.6	124	100.0

Data curation	47	37.9	54	43.5	12	9.7	9	7.3	2	1.6	124	100.0
Complying with the various mandates of funders	36	29.0	65	52.4	16	12.9	7	5.6	0	0.0	124	100.0
Use of instructional repository	35	28.2	68	54.8	15	12.1	6	4.8	0	0.0	124	100.0

Source: Field data (2017)

The results shown in Table 5.27, reveal that 99(79.9%) of the respondents stated that there was a skills gap in developing metadata schema, whereas 101(81.4%) asserted that there was a skills gap in the use of ICTs tools and equipment. On the other hand, 8(6.5%) and 9(7.3%) of the respondents pointed out that they had less skills in developing metadata schema and use of ICTs tools and equipment respectively. Furthermore, 101(81.4%) of the respondents stated that there were skills gaps in data curation (capture, appraisal, description, preservation, access and reuse) and 101(81.4%) of the respondents had skills gaps in compliance with the various mandates of funders. There were 103(83%) of respondents who stated that they had skills gaps in the use of institutional repository.

5.6.3 Training needs

The respondents were asked to state the area they would choose for training. The results are presented in Table 5.29.

Table 5.29 Training needs (n=124)

Statement	SA		A		U		D		SD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Metadata training	59	47.6	48	38.7	10	8.1	5	4.0	2	1.6	124	100.0
Data curation	56	45.2	47	37.9	15	12.1	5	4.0	1	0.8	124	100.0
Information and Communication Technologies (ICTs)	66	53.2	42	33.9	13	10.5	3	2.4	0	0.0	124	100.0
Legal and Ethical consideration in RDM	25	41.9	57	46.0	10	8.1	5	4.0	0	0.0	124	100.0
Funder requirement for research data	42	33.9	58	46.8	16	12.9	8	6.5	0	0.0	124	100.0
Data management plans	45	36.3	53	42.7	20	16.1	3	2.4	3	2.4	124	100.0
Data sharing	42	33.9	55	44.4	14	11.3	13	10.5	0	0.0	124	100.0
Security and storage of data	43	34.7	62	50.0	12	9.7	7	5.6	0	0.0	124	100.0
Collaborative partnerships	34	27.4	67	54.0	15	12.1	5	4.0	3	2.4	124	100.0

Source: Field data (2017)

The results reveal that for the majority of respondents, the training required was metadata107 (86.3%), data curation 103(83.1%), ICTs 108(87.1%), and funder requirement for research data 100(80.7%). Furthermore 82(87.9%), 98(79.0%) and 97(78.3%) of respondents agreed that they needed training in legal and ethical consideration in RDM, data management plans, and data sharing respectively. With respect to the training needs, 105(84.7%) and 101(81.4%) respondents needed to be trained in security and storage of data and collaborative partnerships respectively. However, 7(5.6%) and 8(6.4%) of the respondents disagreed that they did not need training on security and storage of data, and collaborative partnership respectively. One of the heads of research [HR12] from Research institute D asserted:

> *"Every researcher must be a data manager with knowledge in statistics; advanced statistics, research methods, searching and retrieval, security, legal and ethical issues, ICTs, quality assurance and new emerging technologies like metadata, RDM software and others. Training never ends must always refresh knowledge on new RDM aspects. We give refresher trainings majorly on data management mainly analysis and use of ICT with the help of a specialist,.....skills to convert, clean and analyze data, data literacy. I think it is inevitable, the institute should establish RDM unit and recruit or train existing staff on knowledge and skills pertaining RDM"*

Furthermore, the respondents who were interviewed also said that RDM refresher courses were necessary in the institutes. This is necessary in servicing and refreshing the staff in up-scaling and rescaling of knowledge, especially in RDM.

5.6.4 Satisfaction with quality of research data services

Table 5.30, show responses on the quality of the research data services offered by library and IT department.

Table 5.30 Satisfaction with quality of research data services (n=124)

Statement	VS		S		N		D		VD		Total	
	f	%	f	%	f	%	f	%	f	%	f	%
Directly participating with librarian and IT specialist on RDM	28	22.6	63	50.8	14	11.3	17	13.7	2	1.6	124	100.0
Providing reference and advisory support to researchers on RDM	23	18.5	55	44.4	11	8.9	24	19.4	11	8.9	124	100.0
Creating web guides and finding aids for data repositories	35	28.2	34	27.4	24	19.4	18	14.5	13	10.5	124	100.0
Providing technical support for research data service systems	22	17.7	51	41.1	16	12.9	18	14.5	17	13.7	124	100.0
Provision of RDM literacy programs	17	13.7	34	27.4	27	21.8	23	18.5	23	18.5	124	100.0
Coordinating RDM advocacy programs between the researchers and the stakeholders	19	15.3	39	31.5	32	25.8	16	12.9	18	14.5	124	100.0

Source: Field data (2017)

The results reveal that 91(73.4%) of the respondents stated that they were satisfied with direct participation with librarian and IT specialist in RDM whereas 78(62.9%) stated that they were satisfied with the services related to the provision of reference and advisory support to researchers on RDM. The findings also illustrate that the respondents were satisfied with the services related to the creation of web guides and finding aids for data reposition as stated by 69(55.6%) of the respondents. However, 31(25%) of the respondents were dissatisfied with quality of services related to the creation of web guides and finding aids for data reposition. Moreover, 73(58.8%) of the respondents were satisfied with the provision of technical support for research data services systems (repository, web portal, creating metadata, RDM system, access and discovery systems), whereas 35(28.2%) were dissatisfied with the quality of services related to technical support for research data services systems. Concerning the provision of RDM literacy programs, the study established that 51(41.1 %) of the respondents were satisfied while 46(37.0%) were dissatisfied with the quality of service related to provision of RDM literacy programs. Further, 58(46.8%) were satisfied with the coordination of RDM advocacy programs between the researchers and the stakeholders while 34(27.4%) of the respondents were dissatisfied with the quality of service related to the coordination of RDM advocacy programs between the researchers and the stakeholders.

5.6.5 Data literacy programs

The era of e-science demands new skill sets and competencies of researchers to ensure their work is accessible, discoverable and reusable. Librarians and IT specialists are well positioned to assist in data literacy education as part of their liaison and information literacy services. The results in Figure 5.8 show responses on data literacy programs.

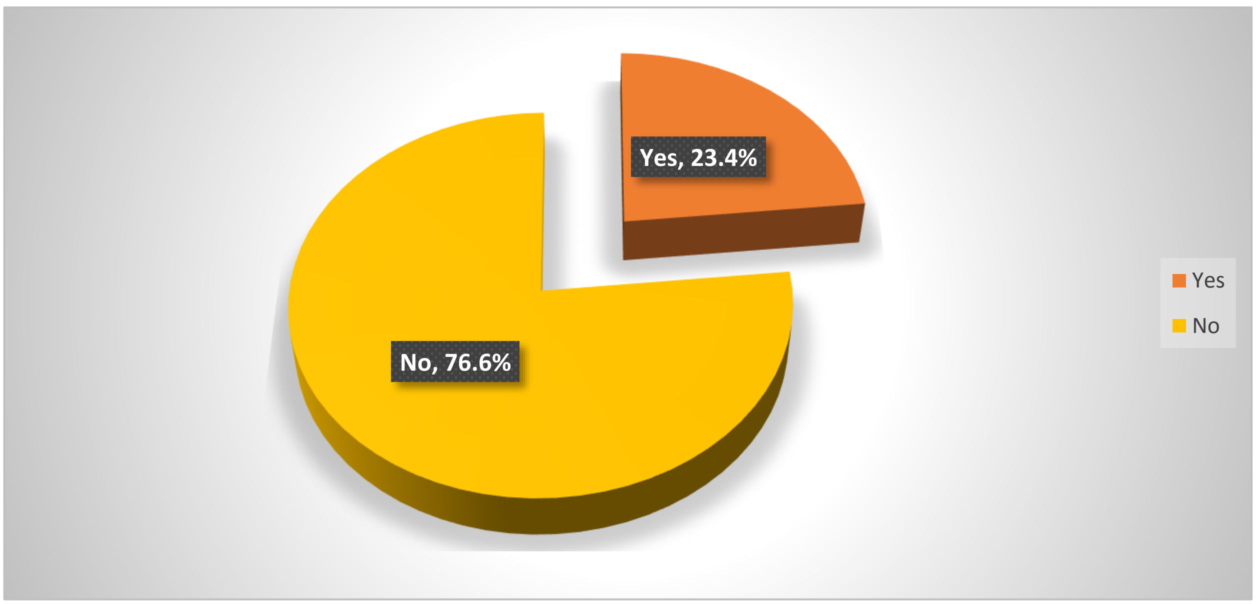

Figure 5.8 Data literacy programs (n=124)

The results in Figure 5.9, reveal that majority of respondents 95(76.6%) had never undertaken any research data literacy program in the research institute where they were working and 29 (23.4%) agreed that they had undergone literacy programs. The data literacy programs undertaken by few respondents were organized by international collaborating partners in the form of workshops and seminars, in-house training on statistical packages, and RDM training by World Agroforestry Centre (ICRAF).

5.7 Level of ICT preparedness in RDM

According to Arzberge et al. (2014), technical infrastructure must be robust in terms of long term and diverse use, flexible to respond to the continuous and rapid changes in RDM which could facilitate data capture, appraisal, description, preservation, access, reuse and sharing. DCC lifecycle model (Higgin, 2008) underpins the ICT preparedness variable because it builds frameworks of standards and technologies in every stage of data curation lifecycle. In addition, digital research data, by its very nature is susceptible to technological changes from the moment of creation. In the entire lifecycle of data curation, ICT infrastructure is necessary for effective and robust RDM. The CCM framework (Lyon et al, 2012) informs the ICT preparedness variable by addressing the technical infrastructure capabilities to perform data-driven research which contribute a lot to RDM. Technical infrastructure capabilities should meet user requirements, support data capture, appraisal, description, preservation, access, reuse and sharing of research data. The Cronbach's Alpha values for this question were 0.703 for items in the researcher's questionnaire. This suggested a high internal validity of the test items. The study sought to establish the level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse of research data. Questions E (1-3) of the questionnaire for researchers (Appendix 4), questions E (1-4) of the interview schedule for directors of institute (Appendix 1), questions E (1-4) of the interview schedule for heads of research (Appendix 2), and questions E (1-4) of the interview schedule for heads of IT and librarians (Appendix 3), addressed this research question. The results are presented in the following sections 5.7.1, 5.7.2 and 5.7.3.

5.7.1 ICT tools and equipment

The study sought to establish the ICT tools and equipment used by researchers in the agricultural research institutes. Table 5.31 present the results.

Table 5.31 ICT tools and equipment used by researchers in data curation

Data curation activities	ICT tools and equipment
Research data capture	Lap-tops, cameras, audio recorders, video recorders, tables, laboratory equipment and magnifying glasses, spreadsheet, scanners, electronic weighing balance, GIS
Research data appraisal	RDM policy and ISO standards available in the institute's web portal
Research data description	RDM software – Grain global, statistical packages
Research data preservation	Institute's repository, institute's server, hard disk, external storage,

Research data access	Publications online, web portals, emails, institutes repository, web side, intranet
Research data sharing	Email, publications online, intranet

Source: Field data (2017)

The respondents who were interviewed were asked to state the range of ICT(s) tools and equipments available for data capture, appraisal, description, preservation and access in their research institutes. The results indicate that computers were very common among the institutes whereby each researcher and technician had their own computers, particularly a lap-top. In some institutes, researchers were using tablets, scanners, cameras, mobile apps and GIS to assist them capture, store, manipulate, analyze, manage and disseminate research data or information in real time. ICT tools and equipment have facilitated heads of research and research group(s) in capturing, apprising, describing and preserving quality research data. Similarly, ICTs tools and equipment have aided in gauging if researcher(s) are conducting incorrect research so that they can be corrected appropriately and early on in their research. The level of ICT(s) infrastructure preparedness with regard to RDM was rated between 50% and 80%.

5.7.2 RDM software(s) used

The section addresses the research question: *What type of RDM software(s) do you use in research data capture, appraisal, description, preservation, access, reuse and sharing in your research institute?*

Agricultural research institutes have adopted ICT to a great extent as indicated by the results presented in Table 5.31. It was established that different agricultural research institutes have different software for data curation because of the diverse research being done. Most of the software packages used is more statistical for example SPSS, SAS and Genstart. However, Research institute E was the only institute that had acquired RDM software called Grain global which is inclusive of research data services like web portal, repository, and intranet among others to assist in managing data. The interviews concurred with the same sentiments that researchers gave.

5.7.3 Frequently used data security methods

The respondents were asked to state the frequently used data security methods in their institutes. The results are presented in Table 5.9.

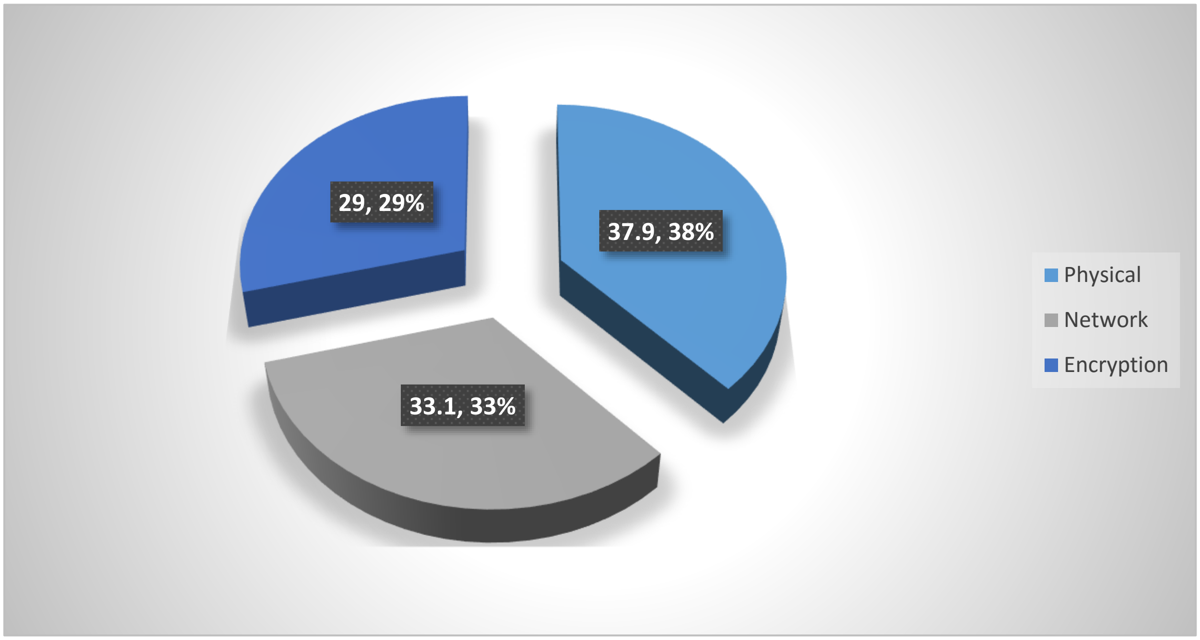

Figure 5.9 Frequently used data security methods (n=124)

The results shown in Figure 5.10, revealed that 47(37.9%) of the respondents stated that the physical security was the most frequently used data security whereas 41(33.1%) stated that they used network security. There were 36(29%) who stated that they used encryption for data security.

The results obtained in the interviews established that sensitive and classified information were kept in a secure place under tight security. It was also pointed out that agricultural research institutes sometimes secured research data by using unique passwords, a centralized server, institute repositories, multiple hard drives and cloud for back up. In Research institute C, the head of research [HR18] said:

> *"There is a records office in every agricultural research institute which preserved research data, There is also a centralized server located in Nairobi and Njoro. The centralized serve is a backup for all the data collected in the field from all the research institutes and the reason for separation is due to security concerns. Also hard copies of research data are kept in secure rooms…security is of paramount importance to the institute because research data in the intellectual assert not only*

to the institute but also to the government and the researcher…..unauthorized persons are denied access and use of password is emphasized in the institute"

5.8 Collaborative partnerships in RDM

Collaborative partnership is needed to enhance access, sharing and reuse of research data. Moreover, collaborative partnership helps in future research, strengthening of the research environment, improving research data workflows and visibility of research outputs (the University of Sheffield, 2016). The Community Capability Model (CCM) framework (Lyon et al., 2012) was used to underpin collaborative partnership variables because it focused on collaborative partnerships in departmental, within the institute, outside the institute, national level, as well as in an international arena. On the other hand, the DCC lifecycle model (Higgins, 2008) outlines access, reuse and sharing of research data and this can only be achieved and utilized through collaborative partnerships. Higgins (2008) posits that while the model provides a high-level view, it can be used in conjunction with relevant reference models for example CCM framework to enhance RDM.

Consequently, collaborative partnerships are a critical factor in growing research through access, sharing and reuse of research data during collaborative projects with inter-state and international partners. The Cronbach's Alpha values for this question were 0.762 for items in the researcher's questionnaire. This suggested a high internal validity of the test items. The research question sought to establish how collaborative partnerships influenced research data capture, appraisal, description, preservation, access and reuse in Kenya's agricultural research institutes. Questions F (1-2) of the questionnaire for researchers (Appendix 4), questions F (1-9) of the interview schedule for directors of institute (Appendix 1), questions F (1-5) of the interview schedule for heads of research (Appendix 2), and questions F (1-4) of the interview schedule for heads of IT and librarians (Appendix 3) addressed this research question.

The parties for collaboration were: collaboration within the discipline/sector; collaboration and interactions across discipline; collaboration with the public; partnership: geographical scale of funding for research; and public-private partnerships. There were a total of 14 items measuring this variable. The researchers who participated in this study were asked to rate the items in this section according to the extent to which they were satisfied or dissatisfied with them. The responses were coded as 1 = Very satisfied, 2 = satisfied, 3 = Neutral, 4= dissatisfied, and 5 = very dissatisfied. While scoring the questionnaires the highest possible score for each item on the Likert scale was 1.0 points and the lowest was 5.0. The highest possible mean score for a respondent was 1.0 and the lowest was 5.0. The midpoint was taken to be 3.0 and this was used to categorize responses as either "satisfied" or "dissatisfied". For each item a mean and standard deviation were calculated. The results are presented in Table 5.32.

Table 5.32 Collaborative partnerships (n=124)

Statement	Mean	Standard deviation
Collaboration within the discipline/sector		
How satisfied are you with collaboration with departmental research groups	2.0968	.75890
How satisfied are you with collaboration across research groups between organizations	2.1694	.72943
How satisfied are you with collaborations organized at national level	2.6371	1.01481
How satisfied are you with international collaborations and consortia	2.2984	.82620
Collaboration and interaction across discipline		
How satisfied are you with individual researchers occasionally collaborating outside their discipline	2.3952	.92674
How satisfied are you with interactions across discipline collaborating through joint conferences or publications	2.2339	.92928
How satisfied are you with formal collaboration between research groups from different disciplines	2.3145	.91391
Collaboration with the public		
How satisfied are you with informational or participative media programs organized to engage the public	2.2177	1.06352
Partnership: Geographical scale of funding for research		
How satisfied are you with research funded through grants from regional agencies	2.0968	.89642
How satisfied are you with research funded by the government	3.0806	1.15186
How satisfied are you with research funding by international bodies	2.2500	.94245
Public-private partnerships		

How satisfied are you with informal partnerships with public and private industry but no funding involved	2.7339	1.02892
How satisfied are you with research co-funded by private industry	2.3468	1.05956
How satisfied are you with established formal co-investment partnership running long term multi-phase research	2.1532	.81694

Source: Field data (2017)

The responses presented in Table 5.31 show that majority of the respondents were satisfied with collaboration with departmental research groups (Mean=2.0968, SD=0.75890), collaboration across research group between organizations (Mean=2.1694, SD=0.72943) and the international collaborations (Mean=2.2984, SD=0.82620). Conversely, respondents were dissatisfied with collaborations organized at national level (Mean=2.6371, SD=1.01481). The study also established that the respondents were satisfied with individual researchers occasionally collaborating outside their discipline (Mean=2.3952, SD=0.92674), interaction across disciplines collaborating through joint conferences or publications (Mean=2.2339, SD=0.92928) and formal collaboration between research groups from different disciplines (Mean=2.3145, SD=0.91391).

The findings also show that the respondents were satisfied with informational or participative media programs organized to engage the public (Mean=2.2177, SD=1.06352). Concerning the level of satisfaction with the geographical scale of funding for research, the respondents were satisfied with research funded internally or through grants from regional agencies (Mean=2.0968, SD=0.89642), and funding by international bodies (Mean=2.2500, SD=0.94245). However, respondents were dissatisfied with research funded by government (Mean=3.0806, SD=1.15186).

Majority of the respondents were satisfied with informal partnership with public and private industry but no funding involved (Mean=2.7339, SD=1.02892), research co-funded by public-private industry (Mean=2.3468, SD=1.05956) and established formal co-investment partnerships running long-term multi-phase research (Mean=2.1532, SD=0.81694).

The interview results show that the heads of research stated that there were collaborative partnerships through trade fairs, where information was shared. The collaborative partnerships in Kenya's agricultural research institute in relation to RDM would vary in terms of access, sharing and collaborations in RDM. Some collaborative partnerships have all rights; some have partial rights over research data based on the memorandum of understanding (MOU) for capturing, description, access and sharing of data. For example, the interview results revealed that one of the agricultural research institutes where data was collected had collaborative partnerships with JKUAT, UON and MoA in research data capture, access, re-using and sharing.

The respondents were asked to state the benefits that accrue from collaborative partnerships. The heads of IT and librarians stated that through collaborative partnerships, they are able to get information easily and it makes data more valuable. It also helps in sharing some sophisticated equipment and skills for data capture, appraisal, description, preservation, data literacy, metadata, access or sharing which the institutes may not have. Through collaborative partnerships, joint publications, networking, joint innovation and patenting is made possible. Furthermore, the heads of research interviewed said that collaborative partnerships help to build and empower those without resources. It also enhances the sharing of statistical data and learning of new skills.

It was also necessary to determine the major challenges the institutes face with regard to collaborative partnerships. Majority of the respondents interviewed stated that there was mistrust with regard to collected data; whether the data collected was inaccurate or correct. One of the heads of research [HR8] said:

> *"Some donors disappear with data. Donors who control finances run the show and at times research delays due to late releasing of finances. Some of them even dictate and misuse us…making data to be vulnerable and not accessible to us….as equal partners."*

There was also a challenge associated with the rights to access and publish joint projects and how to share benefits accrued from projects; who is entitled to what and what proportion. Further, one institute director [D1] interviewed said:

> *"Some organizations refuse to share and access data……other are even worse, they don't include the names of researchers who participated in the research project when they write publication and even acknowledge them. In some instances some make inferences that give negative publicity and can damage the name of the research institute"*.

Based on the results, it is clear that the majority of the respondents agreed that there were more benefits on collaborative partnerships despite the challenges.

5.8.1 Data management plan

The respondents were also asked whether they had a data management plan. The results are presented in Fig 5.10.

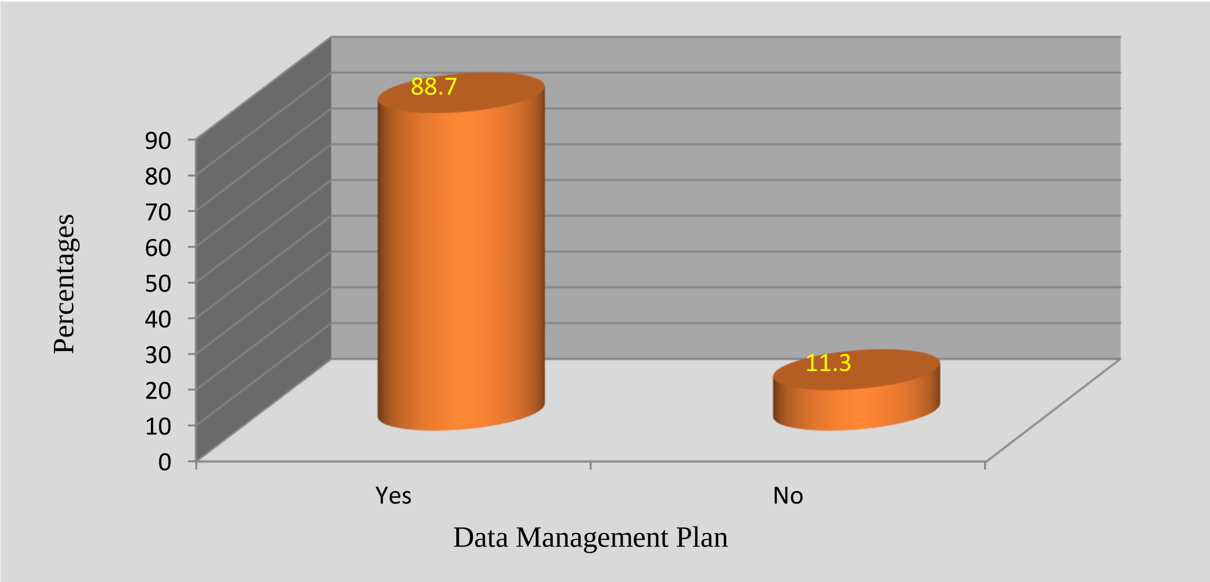

Figure 5.10 Data Management Plan (n=124)

The results in Figure 5.11 indicate that 110(88.7%) of the respondents had a data management plan referred to as SOPs whereas 14(11.3%) didn't have a data management plan.

5.9 Benefits, challenges and recommendations of RDM

The study sought to establish the benefits, challenges and recommendation of RDM from the respondents in the research institutes. The results are presented in the following sections 5.9.1, 5.9.2 and 5.9.3.

5.9.1 Benefits of RDM

The study sought to establish the benefits of RDM. The findings are indicted in Table 5.33.

Table 5.33 Benefits of RDM

Respondents	Responses from the questionnaire
R6	*Referencing purposes, re-using and sharing research data, high quality of data generated and preserved*
R97	*RDM enhances research innovation, enhance the use of technology in research like RDM software linking with intranets, web portal, and others*
R103	*Enhance data mining, avoid duplication hence reduce the cost doing research; security of research data is boosted.*
R66	*Increases the ability to do searching and retrieval, preservation of data will be paramount of which will assist in referencing*

Source: Field data (2017)

The results from interviews also gave similar responses and their responses are summarized in words of two respondents, one is a/the librarian [L1] and the other is the head of research [HR18]. According to the testimony of [L1]:

> *1st it brings about Provision of quality services / reports and information.*
>
> *It also helps Standardization of data / enhances utilization of data and the quality of reports.*
>
> *It also brings about facts based on research*
>
> *It also enhances utilization of knowledge which boosts innovation.*
>
> *It also brings about trust and improves the image of the organization since you can relate yourself with information.*
>
> *Library and IT departments will play a crucial in RDM*

According to the opinion of [HR8]:

> *"Data will be available to researchers/scientists and they can access and share it, use modern technology in data curation. Another benefit is the new findings after re-analysis of existing data. Systematic way of handling research data from its creation to access and sharing"*

5.9.2 Challenges encountered in RDM

The challenges of RDM in agricultural research institutes is summarized on table 5.34

Table 5.34 Challenges encountered in RDM

Respondents	Responses
R7	*Lack of funding hinders the establishment of RDM unit in the institute, Lack of legal framework to guide RDM practice, No centralized RDM policy from the headquarters to guide other institutes.*
R46	*Untrained staff especially researchers, librarians, IT staff and RDM stakeholder on RDM, lack of awareness on RDM, RDM literacy and advocacy program*
R87	*There is limited and poor quality ICT tools, equipment and services, inadequate and unqualified personnel to handle technical infrastructure, missing out on RDM software for all Kenya's agricultural research institutes*
R119	*Poor storage, access and sharing of information especially if a researcher/scientist leaves the institute, comprehensive and coherent RDM policy and regulations are lacking.*

Source: Field data (2017)

Majority of the respondents interviewed cited the following as challenges encountered in RDM

- *"The most pressing challenge is lack of RDM policies- this has hampered development in the RDM unit within the institute".*
- *"There is also a lack of proper ICT infrastructure setup since the machines needed are very expensive"*
- *"Lack of information by the management on the importance of RDM, some don't understand why the organization should use a lot of money on RDM… institutes should embrace Training/ sensitization of staff on RDM issue".*
- *"There is also the lack of funding which has hampered researchers in doing research in order to have research data to manage and also procurement of equipment"*

5.9.3 Recommendations to mitigate the challenges

Following the cited challenges, the respondents were asked through interviews and questionnaire to suggest recommendation to mitigate the challenges. The respondents proposed the following recommendations: Need for government support through enactment of RDM legal framework and funding; technical infrastructure development; capacity building; formulation of RDM policies and regulation; data literacy and advocacy programs be established; establishment of RDM unit to oversee RDM; and enhance collaborative partnership in RDM. Of the proposed solutions, enactment of RDM legal framework, capacity building and technical infrastructure development were cited by the majority as the core determinants of RDM in agricultural research institute and should be facilitated. In this regard, legal framework is the guiding principle to the RDM activities while capacity building and technical infrastructure development are the critical pillars and functionalities in RDM.

Summary of Interpretation and discussion of findings

The chapter discussed about data analysis and presentation of the findings based on the themes draw from the research questions of the study. The findings revealed that RDM is considered imperative in enhancing research in Kenya's agricultural research institutes.

However, the findings showed that lack of RDM legal framework is a major impediment to RDM in all agricultural research institutes because the recent KALRO Act (No.17 of 2013) does not clearly define how research data generated should be management, accessed, shared and reused. The finding further establishes that research data from capturing, appraisal, description, preservation, access and reuse is still faced with many challenges including: Inadequate policies and regulation which are comprehensive and coherent for every stage in data curation, and insufficient modern technical infrastructure compatible to repositories, metadata, security systems, searching retrieval mechanisms and other data management systems. In addition, ownership of research data, person/s solely responsible for RDM, preservation, access and sharing of research data was found to be a problematic issue.

Further still, the finding revealed that the agricultural research institutes had inadequately trained RDM staff and the few were rarely supported to attend conferences, workshops and seminars to enhance their knowledge and skills in RDM. Likewise, librarians and IT staff are not considered to play any role in RDM, furthermore data literacy and advocacy programs were lacking. In the same vein, collaborative partnerships lack a uniform working model for research data capturing, appraisal, description, preservation, access, reuse and sharing in all Kenya's agricultural research institutes, this makes it difficult to construct a framework from which to draw lessons and experience. Government support of RDM is minimal or none existing through budgetary allocation for RDM functions and activities which could have accorder RDM the status of a department in every agricultural research institute.

Though the findings showed that KALRO is currently restructuring and coordinating agricultural research institutes, the vast majority of them (agricultural research institute) are reconsidering establishing RDM unit to coordinated research data in various department for access, reuse and sharing. Nevertheless, proposed intervention to the challenges were suggested which could improve management, sharing, and reuse of agricultural research output.

The finding were interpreted and discussed in chapter six (6) based on the research questions of the study.

CHAPTER SIX

INTERPRETATION AND DISCUSSION OF FINDINGS

6.1 Introduction

This chapter presents the discussion and interpretation of findings obtained from qualitative and quantitative data collected and analyzed in chapter five (5). The interpretation and discussion of findings is an important part of the study as it describes the relevance of results in relation to the research question being investigated. According to Fain (2013:272), discussion chapter involves organizing and explaining the meaning of the findings gathered from both quantitative and qualitative methods. In the same vein, interpretation begins with an attempt to explain the research findings within the context of the theoretical framework and prior empirical knowledge. Kothari (2004) in this regard concurs that interpretation and discussion of findings provides a theoretical basis for further research. LoBiondo-Wood and Haber (2014) corroborate with Fain (2013) that in an interpretation and discussion chapter, the researcher brings data to life by deducing and discussing the results.

The study sought to examine Research Data Management in Kenya's agricultural research institutes with the view to proposing interventions to improve management, sharing and reuse of agricultural research output. The following research objectives were addressed: to assess the status of research data management in Kenya's agricultural research institutes; and determine the legal and policy framework, ICT infrastructure and human capital that is needed or available to facilitate RDM in Kenya's agricultural research institutes. The interpretation and discussion of the findings were based on the research questions restated as follows:

1. How does the availability or absence of legal, policy and regulations affect the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes?
2. How do Kenya's agricultural research institutes capture, appraise, describe, preserve, and make accessible for reuse its research data?
3. What knowledge, skills and training are needed to capture, appraise, describe, preserve, and make accessible for reuse its research data?
4. What is the level of ICT preparedness in Kenya's agricultural research institutes for the capture, appraisal, description, preservation, access and reuse of research data?

5. How do collaborative partnerships influence the capture, appraisal, description, preservation, access and reuse of research data in Kenya's agricultural research institutes?

The respondents comprised directors, heads of research, researchers, heads of IT and librarians from six (6) agricultural research institutes namely: Research Institute A, Research Institute B, Research Institute C, Research Institute D, Research Institute E, and Research Institute F. The study was underpinned by Data Curation Centre (DCC) lifecycle model (Higgins, 2008) and Community Capability Model (CCM) framework (Lyon et al., 2012). Furthermore, the study adapted pragmatism ontology with a focus on mixed methods epistemology that enabled the researcher to collect quantitative data from a large sample of researchers from the agricultural research institutes in Kenya and also collect qualitative data from directors of institutes, heads of research, heads of IT and librarians through interviews. A Similar approach was adopted by Shakeri (2013) in a study on data curation perspectives and practices of researchers at Kent State University, UK.

6.2 Background information of the respondents

Background information was collected on research institute of affiliation of respondents, gender, age, field of specialization, years of working experience and highest academic qualification of the respondents. Data Curation Centre (DCC) lifecycle model (Higgins, 2008) and Community Capability Model (CCM) framework (Lyon et al., 2012) that are used to underpin this study illuminates on the background of the respondents by focusing on the following: as training tools for data creators, data curators and data users; to organize and plan their resources; and to help institutes identify risks to their digital assets and plan management strategies for their successful data curation. In the same vein, the mentioned theories are tools developed to assist the research community in growing their capability in areas such as ICT infrastructure; human capacity; legal, policy and regulatory framework; and collaborative partnerships to perform data-intensive research.

The results revealed that the majority of the respondents 35(28.2%) were working in Research Institute A, with the main research focus on food crops like cereals, grain legumes, and root and tuber crops, while 23(18.5%) were working in Research Institute F focusing its research on coffee along the value chain. There were 21(16.9%) who were working in Research Institute C with the mandate to conduct basic, applied and adaptive dairy research, 16(12.9%) were working in

Research Institute B with the main focus in conducting research in tea products whereas 15(12.1%) were working in Research Institute E with the mandate to conduct research on conserving plant, animal and microbial genetic resources. Only 14(11.3%) were working in Research Institute D which focuses on conducting research in laboratory-based biotechnology. According to KALRO (2016), Kenya currently has 16 agricultural research institutes established under the KALRO Act of 2013. The 16 agricultural research institutes conduct research in various agricultural areas such as: food crop, livestock, biotechnology, genetic resource, poultry, sugar, tea, coffee, agricultural mechanization, apiculture, horticulture, and industrial crops. Each research institute has a national mandate for specific major commodities with the objective of disseminating appropriate information and technologies to intended users.

The findings further revealed that 68(54.8%) of the researchers were male while 56(45.2%) were female. These results suggest that almost an equal number of the researchers from either sex worked in the research institutes and participated in this study. It was also noted that 30(24.2%) of the researchers were aged 36-40 years, while 23(18.5%) were above 50 years old. Another 19(15.3%) were 31-35 years, 41-45 years, 46-49 years old respectively. There were 14(11.3 %) who were 26-30 years old. This implies that the majority of respondents were between 26-50 years old with knowledge and experience in research and management. A survey conducted by Beintema and Stads (2017) on investment and human resource capacity in African agricultural research found that 70% of PhD qualified agricultural researchers were over 50 years old and this posed a big retention threat to research institutes because this workforce was likely to retire in the next ten years leaving the institutes without senior researchers and mentors to train junior staff. The institutes needed to aggressively develop succession strategies and training programs to ameliorate the negative impact that would arise from retirements and other factors that would cause staff turnover.

The results revealed that 10(8.1%) of the participants had their field of specialization in records management, 17(13.7%) specialized in agricultural pathology whereas 14(11.3%) had their specialization in agricultural economics. Majority of the respondents specialized in Biochemistry 24(19.4%), followed up by Breeding at 20(16.1%). The findings corroborate the findings obtained from interviews where most respondents confirmed that the most recruited cadre of researchers were breeders, agricultural pathologists, biochemistry and agricultural economists because they were the most required cadre of staff in every research institute. This result is in tandem with the assertion of Maru (2004); and Beintema and Stads (2017) who pointed out that in order to fulfill research mandates effectively, it is important for agricultural research systems to have a well-balanced pool of researchers not only in terms of qualification levels, age distribution, and gender, but also in terms of research discipline. Furthermore, the results revealed that 44(35.5%) of the respondents were bachelor's degree holders and 41(33.1%) were masters' degree holders. Another 21(16.9%) were PhD degree holders whereas 11(8.9%) were holders of higher diploma qualification and only 7(5.6%) were holders of diploma qualification. The findings seem to suggest that agricultural research institutes in Kenya recruited qualified researchers with degree qualifications. Beintema and Stads (2017) explains that agricultural research institutes should recruit BSc or MSc qualified researchers with minimum number of PhD-qualified research scientists necessary for the conception, execution, and management of high-quality research and for effective communication with policy makers, donors and other stakeholders.

6.3 Legal, policy and regulatory framework for RDM

Agricultural research institutes are increasingly getting involved in data-intensive research projects that cut across disciplinary borders and also involve communities of researchers participating in large-scale collaborations. In this respect, there is a need to develop legal frameworks, policies and regulations on RDM to facilitate systems and services in enabling research data to be managed, accessed, shared, reused and secured. Fitzgerald and Pappalardo (2007) assert that RDM occurs in legal and policy contexts and the principal areas of law that cover RDM includes copyright, moral rights, patents, confidentially, contract, and privacy. Erway (2013); and Higman and Pinfield (2015) point out that failure to establish legal and policy protocols for RDM is likely to diminish the potential for valuable research outputs to be made available for access, sharing and reuse.

6.3.1 Legal framework governing RDM

The findings of the study revealed that there was no legal framework governing the RDM in Kenya's agricultural research institutes despite the fact that they recognized that it was vital in the administration of RDM. This was supported by 90(72.6%) of the respondents who agreed that there was no legal framework and 10 (27.4%) disagreed as illustrated in Figure 5.1. The same sentiments were echoed by the directors of institutes, heads of research, heads of IT and librarians in their interviews. A possible explanation for these results may be the lack of adequate legal framework attached to the KALRO Act (No.17 of 2013). In essence, the establishment of KALRO

Act (No.17of 2013) to coordinate agricultural research in the country did not clearly define how research data generated in the research institutes should be managed to ensure the continuous preservation, access, sharing and reuse of the data. This is in agreement with the assertions by Pinfield, Cox and Smith (2014) in their study on RDM and libraries that the key components of RDM in every research institution is the legal framework specifying how the strategies are to be operationalised through regular procedures and a set of complementary policy frameworks covering access, sharing, intellectual property rights and reuse.

In tandem with the provision of the CCM framework (Lyon et al., 2012), legal frameworks should be comprehensive, articulate, coherent and consistent with all stages of data curation lifecycle including human resource capability, technical infrastructure and collaborative partnerships. The CCM framework points out that the absence of a well-built RDM legal framework suggests a limited or uncoordinated response and obligations to legal, policy and regulatory issues by the organizations. The results revealed that where basic legal frameworks existed they were disjointed, researchers were not aware of them; the corresponding procedures to the legal frameworks were not well enforced and therefore hindered RDM. The CCM framework asserts that RDM and the law have a bearing on how research institutions respond to data curation, opportunities presented by technical infrastructure, a broad range of human capabilities and policy guidelines to ensure research data is managed, accessed, shared, and reused by the research community (McGeever, Whyte and Molloy, 2015).

6.3.2 Policies and regulations available to facilitate RDM

The study sought to determine whether there were policies and regulations governing RDM. The results established that there were policy and regulations that govern research data management practices in a majority of the research institutes where the study was done (Mean=2.0806, SD=0.69372). A possible explanation for this might be that Kenya’s agricultural research institutes had the mandate to formulate policies and regulations that govern RDM and ownership of data. Anderson (2004); and National Institutes of Health (NIH) (2003b) observed that policy and regulations for RDM cover a wide range of areas that relate to data curation, sharing, reuse, intellectual property, ethical requirement, ownership of data, security, human resource capability, technical infrastructure capability and collaborative partnerships. The results established that as much as the RDM policy and regulations were available at Kenya’s agricultural institutes, their application in RDM practice was questionable because there was lack of RDM unit to enforce the policies and regulations. The results from interviews revealed that departments handled their research data independently of their other departments making coordination of RDM and its policies a challenge. The findings revealed absence of coordination of RDM portfolio among departments in the institutes surveyed. This calls for the establishment of RDM unit to perform the coordination function. Qin (2013), in a study on infrastructure, standards, and policies for RDM in Syracuse University found that RDM policies and regulations at national and institutional levels establish the framework for individual researchers and departments to work collaboratively.

6.3.2.1 Role of policies and regulations in research data lifecycle

The respondents were asked if the agricultural research institute had RDM policies and regulations governing research data capture, appraisal, description, preservation, access, use and reuse, and sharing. The general findings revealed that agricultural research institutes had RDM policies and regulations in the following stages of data curation: research data capture (Mean=2.0000, SD=0.79633), appraisal (Mean=2.3065, SD=0.97261), description (Mean=2.0726, SD=0.78785), preservation (Mean=2.0403, SD=0.84005), access (Mean=2.0645, SD=0.83366), use and reuse (Mean=2.0887, SD=0.79646), and sharing (Mean=2.1210, SD=0.87970). It emerged that though policies and regulations do exist they were not applied effectively at every stage in the research lifecycle as envisaged by the DCC lifecycle model (Higgins, 2008). It was revealed that if policies and regulations were applied compressively and consistently at every stage of the lifecycle then efficacy of RDM could be felt in agricultural research institutes. As pointed out already, though the policies were available in the research institutes their applicability remained a challenge. Erway (2013); and Jones, Pryor and Whyte (2013) emphasize that policies and regulations at every stage of research lifecycle ensure that consistent data management standards and quality are maintained in order to foster management of and access to research institute's intellectual assets and also provide uniform requirements to facilitate data understandability and sharing among research data stakeholders.

These findings corroborate with those obtained through interviews, where most respondents confirmed that though policies and regulations covered every phase of research data lifecycle at agricultural research institutes, there was a lack of coordination in their applicability. This situation meant that every researcher, heads of research or departments managed their own research data independently without reference to standards, policies and regulations. National University of Singapore (2017) in a study on data management policy found that policies were necessary in providing clarity on what is expected by the research institutions and where responsibility for RDM resided. Moreover, coherent vision expressed through a policy is useful in coordinating the broad range of interests. Besides, RDM policies provide a framework for an overarching governance to avoid application of policy unevenly.

The findings revealed that other stages of research data lifecycle such as appraisal, description and preservation did not have adequate and comprehensive policies and regulations governing them, making RDM in these areas inconsistent and incoherent. This was exacerbated by the lack of expertise in these areas. Mossink, Bijsterbosch and Nortier (2013); and MacKenzie (2014) in studies covering a range of RDM policies based on DCC lifecycle model found that whereas most research institutes had policies covering RDM, there was inadequacy of policies covering appraisal, description, preservation, security, and sharing. They also found that while most research institutions tended to have overarching RDM policies they lacked specific policies governing every phase of RDM lifecycle. Similarly, a study done by Bracke (2011) on curation of agricultural data established that there were often no data standards within a discipline and researchers often had very idiosyncratic research practices which increased complexity in the agricultural research data.

6.3.2.2 Policies and regulations on RDM capabilities

The findings revealed that in the agricultural research institutes surveyed, there were no clear policies governing knowledge, skills and training (Mean=2.4274, SD=0.01351), technical infrastructure (Mean=2.1290, SD=0.91924) and collaborative partnerships (Mean=2.0403, SD=0.89624) as illustrated in Table 5.3. These findings in particular revealed that there were no policies designated for human resource capability as well as technical infrastructure to facilitate RDM. The policies that were available generally focused on human resource management and technical infrastructure. Mossink Bijsterbosch and Nortier (2013) asserted that research institutions should build a trustworthy RDM unit with qualified human resource capability and technical infrastructure in order to raise awareness among researchers regarding important research data related aspects such as access, mining, sharing, reuse and data service.

The findings seem to concur with other previous studies (Mullins, 2014; Higman and Pinfield, 2015; Erway, 2013; Keralis et al., 2013) which found that many research institutions did not requisite regulations and were now developing new institutional RDM policies with the involvement of stakeholders. Besides, extant RDM policies and regulations on human resource capability, collaborative partnerships and technical infrastructure should be enforced.

6.3.2.3 Policies and regulations on ownership of research data

The findings revealed that after most research institutes were surveyed, the researchers had the right to research data. The results revealed that research institutes surveyed owned the right to research data (Mean=1.7823, SD=0.76041); researchers owned the right to research data they created (Mean=1.7016, SD=0.70974). These findings would appear to suggest that RDM policy and intellectual property policy in the research institutes surveyed clearly articulated the ownership of research data. However, some of the issues emerging from this finding relate to the weak enforcement of the RDM policy and intellectual property concerning ownership of research data and as a result valuable datasets are getting lost or discarded. Another issue that emerged was lack of policy on the responsibility for RDM especially research data left behind by researchers. Erway (2013) noted that most RDM and intellectual property policies tended to be fairly toothless in enforcing compliance and were dysfunctional. Consequently when most of the researchers left or disengaged with the research institutes, they took research data with them or left it behind unattended to. In such cases, the research data got lost or was discarded. Council on Library and Information Resources (2013) suggests that to enforce compliance with RDM, researchers must be compelled to adhere to RDM policies by making compliance a requirement for tenure and promotion. Similarly, Mossink Bijsterbosch and Nortier (2013); Erway (2013); Higman and Pinfield (2015); and Cox and Pinfield (2014) agree that there is a need for policies and guidelines with solid institutional mandate to govern ownership of data created within the agricultural research institutes as this is a problematic area.

CIARD (2012) in a study on ownership of the research output established that research outputs in the form of research data, research reports, technical reports, manuals, and handbooks that were done by researchers at KARI, KEFRI and MoA, belonged to the research institute. Besides, Indian Council of Agricultural Research (ICAR) (2014); and Koopman and De Jager (2015) affirm that all research data collected or generated from research, educational or allied activities conducted at the agricultural research institutes and using of institutes resources shall belong to the institute including research data generated by consultants hired by the research institute. However, ICAR notes that exceptions to ownership may be subjected to approve agreements for collaborative research, sponsored research, contract research or contractual services rendered to third parties. Consequently, ownership of research data must be determined based on existing research institutions RDM policies, procedures and agreements and more importantly researchers must be aware that ownership of research data can also be influenced by funding agreements or collaboration agreements with other institutions (Australian Catholic University, 2017).

6.3.2.4 Quality assurance and control measures during data curation

The findings revealed that majority of the respondents agreed that there was quality assurance and control measures in place to govern data capture, appraisal, description, preservation, access and reuse. The findings further revealed that the respondents agreed that there was quality assurance and control measures in place for data capture (Mean=1.8306, SD=0.69518), appraisal (Mean=2.1290, SD=0.90138), description (Mean=1.9355, SD=0.82385) and preservation (Mean=1.9597, SD=0.85918). Furthermore, the result showed that there was quality assurance and control measures in place for data access (Mean=1.9194, SD=0.73911) and reuse (Mean=1.9677, SD=0.76441) as illustrated in Figure 5.3. These results may be explained by the fact that KALRO has embraced ISO 9001: 2008 and made efforts to become ISO 17025:2005 compliant. These efforts suggest that KALRO is committed to quality service delivery to clients (KALRO Service Charter, 2016). These findings concur with the findings of a study by Marchionini, Lee and Bowden (2012) on curating for quality which emphasized that data creators and curators in agricultural research institutes must identify indicators of quality; develop and use tools and techniques that ensure useful, usable, and accurate metadata discovery, management, and sharing; create and use best practices and open standards whenever possible; and provide auditable validations for quality.

All agricultural research institutes surveyed were ISO compliant and this has enabled them to develop Standard Operating Procedures (SOPs) used in capturing, preserving, accessing, re-using and sharing research data. Van den Eynden et al., (2011) acknowledge that quality control of data is an integral part of all research and takes place at various stages including data capture, appraisal, description, preservation, access and sharing. In this respect, Van den Eynden stresses the importance to assign clear roles and responsibilities for data quality assurance at all stages of data curation and develop suitable procedures for RDM.

6.3.3 Legal and policy challenges facing RDM

Concerning the legal and policy challenges facing agricultural research institutes in the capture, appraisal, description, preservation, access, reuse and sharing of data; directors of institutes, heads of research, heads of IT and librarians who were interviewed, stated that in some cases, it was difficult to monitor the reports or research data from researchers' work or research data in departments because most of the researchers used a personal computer for managing research data. The findings seemed to suggest that there was no centralized coordination of RDM to facilitate access, reuse, and sharing of research data. The finding is similar to that of Jahnke and Asher (2012) that research institutes and researchers were not well positioned to meet the legal, policy and technical challenges of RDM without the support of libraries, information technology units and RDM professionals who possess both technical and research expertise.

The findings from the interview further revealed that plagiarism of research data collected was a big challenge facing researchers in all agricultural research institutes surveyed due to the absence of legal framework and ineffective RDM policy and regulations. Ethics and integrity in the conduct of research are critical to the advancement of scientific knowledge. For this reason Fossey (2007) posits that when research communities and funders are deceived by giving false information or presentation of false results, it is scientific fraud that may be classified as: fabrication, falsification, plagiarism and misappropriation, as the case may be. In this regard, failure to comply with the ethical code of conduct for researchers may be regarded as gross misconduct which leads to disciplinary action. It is therefore imperative that researchers understand the concepts that underlie ordinary morality which directly affects research data and its management. It is imperative for the agricultural institutes to establish a unit for promoting advocacy around legislative, policy, regulatory and ethical matters affecting RDM. Such entities would also be responsible for technical infrastructure, intellectual property administration, and research integrity (Erway, 2013).

6.4 Research data management

Research data are important this includes expensive output of the scholarly research process across all disciplines. Therefore curation of research data manifests itself as being of critical importance. Data curation is the activity of managing and promoting the use of data from its point of creation to ensure it is fit for contemporary purpose and available for discovery and use (Hinnant et al., 2012; Walton, 2010; and Lord et al., 2004). Data curation consists of a range of activities and processes focused on capture, appraisal, description, preservation, access and reuse which adds value to research data throughout its lifecycle. DCC lifecycle model (Higgins, 2008; and Heidorn, 2011) supports data curation by providing a high-level overview of the stages required for successful curation of research data and good RDM. The stages of data curation as presented by DCC lifecycle model (Higgins, 2008) includes; capture, appraisal, description, preservation, access, use and reuse. These stages form the foundation of RDM from which activities of data curation are generated (Palathingal et al., 2015; and Higgins, 2008). These activities enable the CCM framework (Lyon et al., 2012) to define roles and responsibilities, build a framework of standards, and technologies to implement in RDM.

The study sought to establish RDM (research data capture, appraisal, description, preservation, access, reuse and sharing) in agricultural research institutes in Kenya. The findings on these variables are discussed and interpreted in the following sub-sections:

6.4.1 Research data capture

Data capturing is the process of collecting data which will be processed and used later to fulfill certain purposes (Van den Eynden et al, 2011). The findings revealed that the commonly used method of capturing research data by researchers was the use of a questionnaire, interviews, audio recordings, cameras, GIS, laboratory experiment and field experiment. Thereafter, the captured research data is analyzed and disseminated to the researchers or heads of research for consultation, use or sharing purposes. These results may be explained by the fact that different research projects have different methods of data capture. The finding also suggests that research institutes are trying to adopt modern technologies in data collection and were using Standard Operating Procedures (SOPs) as a guiding principle during data collection. Lindquist (2013) in a study on method for data capture found that if older technologies continued to be used in capturing data moving forward, there will be challenges when verifying data and gathering endpoint conclusions because the agricultural sector evolves at a rapid rate and this requires modern ICTs to capture data and process in real time. Statistics Canada (2015) in a study on data collection, capture and coding established that during data collection the following should be taken into consideration: careful planning of the collection process; using appropriate real time data collection method; designing the collection process; establishing appropriate sampling among others. In addition, accurate data collection is essential to maintaining the integrity of research and generating gradable research out.

Concerning the format in which data was generated, the findings established that majority of the respondents generated their research data in images 101(81.5%), spread sheets 98(79.1%), video 84(67.8%), data statistical (SAS, SPSS) 105(84.7%) and database 106(85.5%) as illustrated in figure 5.5. Three quarters of the respondents generated their data in scanned document format, images, web format, computer aided design, geographic information systems and data XML. These results suggest that there were different formats of capturing research data in agricultural research institutes using ICTs. These formats in which research data is collected may not be accessed, reused and shared by other researchers in the research institute or from other research institutes because they could be outdated and their compatibility with modern technology is a challenge. At the same time, preservation methods of different formats are poor hence affecting their durability which in turn affects access to research data.

These findings corroborate those of Peters and Dryden (2011) who in a study on assessing the academic library's role in campus-wide research data management found that 80% of participants generated data with images, scanned documents, spreadsheets, and text formats. Another 20% indicated that all formats are generated at any particular point in the process of generating research data. Van den Eynden et al., (2011) say that the format and software in which research data are created and digitalized depend on how researchers plan to analyze data, availability of software or is determined by discipline-specific standards and customs. Studies done by Bracke (2011); Karasti, Baker and Hakola (2006); Peter and Dryden (2011); and Heidorn (2011) on data curation using DCC lifecycle model contended that capturing of research data plays a critical role in RDM because it influences and determines the quality, comprehensiveness and consistency of appraisal, description, preservation, access, reuse and sharing of research data. United Nations Environmental Program (UNEP) (2015) claims that in developing countries, agricultural research data existing in different formats or software versions scattered across servers, computers, storage devices or other filing systems, which makes it difficult to change into modern formats or versions and in the process research data is lost.

In addition to the methods and formats of capturing research data, the respondents were asked about the amount of data the institute had generated. The general findings revealed that research institute A generated data from 1gigabyte to >500 petabyte, research institute D and E generated the least data from 1gigabyte to >500 petabytes. The large amount of data generated by research institute A may be attributed to the fact that it researches in a wide range of areas such as cereals, grain legumes, and root and tuber crops. On the other hand research institutes D and E are the newly established institutions following the enactment of KALRO Act of 2013. Furthermore, the findings seem to suggest that the huge amount of data generated is scattered in different research institutes, departments, heads of research or individual researchers making access, reuse and sharing a challenge. The KALRO Act (No.17 of 2013) however, does not clearly define how research data generated in the research institutes should be managed to ensure the continued preservation, long-term management, sharing and reuse of agricultural research data.

These findings are not surprising given that generation of research data is depended on the availability of funds. Institutes that are well established have the finances to sponsor research and also receive funding from external agencies. On the other hand, those institutes with limited funding rely on external funding from such agencies as World Bank, FAO, International Livestock Research Institute (ILRI) and other development partners. The findings corroborates with Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data (2016) that the sustainability of RDM represents a challenge within the existing funding structures especially research institutes in developing countries and at the core of this particular challenge are issues related to the eligibility of funding for research.

6.4.2 Appraisal of research data

Appraisal and selection is the process of evaluating research data in order to decide which data to retain over the long term, which to retain for the medium term and which to discard (Higgins, 2012; and Morris, 2013). DCC lifecycle model (Higgins, 2008), states that appraisal is an ongoing process in RDM and documented guidance, policies and legal requirements should be adhered to.

The respondents were asked to state the research data appraisal checklist in their institutes. The findings of the study indicated that uniqueness, repeatability, scientific/historical value, complementary/added value, reuse value, substantiveness, access, volume and cost-effectiveness were the items considered as research data appraisal checklist in their institute. This range of checklist speaks to the value institutes attach to data appraisal.

The respondents were further asked to state the tools used to guide the appraisal of data. The findings disclosed that all the research institutes preferred using research institute's policy. For example 21(17.0%) and 11(8.8%) of respondents from research institute A and B respectively preferred research institute's policy whereas 17(13.7%) and 8(6.4%) of respondents from the same research institutes preferred RDM policy, while appraisal and selection policy was preferred least. Furthermore, the findings seem to suggest that RDM policy and appraisal and selection policy were rarely used by the research institutes perhaps due to inadequate guidance on their use. Attempts have been made to develop a framework for appraising data. In this regard, Harvey (2006) tried to create a framework for appraising and selecting research data by adopting some archival appraisal theories and methods and reviewing existing appraisal/selection policies for preservation and digitization. In addition, Harvey (2006) and Morris (2013) created ten appraisal criteria for data appraisal that included: value, physical condition, resources available, use, social significance, legal rights, format issues, technical issues, policies and documentation. Higgins (2008) with regard to DCC lifecycle model states that appraisal and selection of research data should adhere to documented guidance, policies or legal requirements. Whyte and Wilson (2010); and Tjalsma and Romnouts (2011) affirm that the appraisal and selection policy must fit legal requirements relating to privacy and intellectual property rights, Public Records Acts, national data policies and codes of conduct adopted by the host institution or agricultural research institute or funders.

6.4.3 Description of research data (metadata)

Data description entails derived data; produced research outputs; author publications; data anonymisation; data visualization; data validation; metadata creation; data verification; data interpretation and analysis; and preparing data for preservation (Farnel and Shiri, 2014; Higgins, 2008; and Van den Eynden, 2013). Describing and documenting research data ensures that it can be found or discovered, preserved, accessed, reused, and shared in the long term and this can be achieved by using appropriate administrative, descriptive and other metadata during the time of data creation (Heidorn, 2011; and The University of Queensland, 2017).

There existed a need to establish the respondent's opinions concerning the description of research data. The findings revealed that different researchers adopted different ways of describing their research datasets. For example in research institute A 21(16.9%), 21(16.9%), 25(20.1%) and

20(16.1%) 87(70.2%) of respondents preferred only the description the file names on their hard drives had; using handwritten notes in their lab notebook after the experiments have been completed; describing the data using the column and row labels in their spread sheets; and creating descriptive metadata for each datasets and saving the descriptions with datasets on hard drive respectively. The findings suggest that different researchers adopted different ways of describing their research datasets depending on the research project they were undertaking. There are, however, other possible explanations why results seem to reflect disparities in the way research datasets are described in various research institutes such as the lack of harmonized procedures in data description making accessibility and sharing a challenge. A study by Ball (2012) that explored the feasibility and desirability of a harmonized application profile to improve resource discovery and reuse of research data found that there was a need for comparison of data models and metadata schemes from a variety of disciplines in order to come up with a generalized metadata profile that could be applied to fulfill the requirements of the use cases.

Concerning the creation and capturing of metadata, the findings revealed that 104 (83.9%) of the respondents preferred their data descriptions to be saved in spreadsheets or word processor documents followed by others who preferred creating a title and short textual description for each dataset when submitting to their research institute data repository. Others created rich metadata by recording data at the time of acquisition using a metadata entry form to ensure they did not miss any essential information. Least preferred was the automatic inclusion of instrument metadata in each data file. The findings seem to suggest that the research institutes did not have a uniform way of creating metadata for research data and another possible explanation for this could be the absence of a coordinating unit for metadata creation. This finding resonates with that of Peters and Dryden (2011) who assessed the academic library's role in campus-wide RDM and found that 29.4% of the respondents had no real file or folder naming conventions in place, 5.9% responded they were instructed by the head of research on how to manage the data associated with the project while 11.8% used research institutes standards to describe data but did not specify exactly what those standards were. Furthermore, 29.4% of the respondents used specific file-and folder-naming protocols, 5.9% used the file names generated by the equipment, 11.8% mentioned that data was automatically time-stamped by their equipment and 5.9% mentioned that some of the equipment automatically embedded metadata as data was generated. The respondents further acknowledged that there were no clear methods of data description, thus research data was largely predetermined by experiment design or type of research project and research practice.

Jones, Guy and Pickton (2013) point out that at the capturing stage, file naming, versioning and structuring of files needs to be performed to ensure ease of accessing data when needed, bearing in mind both the short-term and long-term description. Agricultural research data with good metadata attached at the point of capture can expedite data sharing, mining, publishing and citation. Metadata capture is of great value as the more information there is about data, the greater the value of the data, whether automatic or manual and this makes RDM valuable. Moreover, there is a lack of awareness about the importance of metadata among the scientific community and therefore there is a need for research institutions to organize RDM literacy programs to assist researchers and other RDM stakeholders to understand, prepare and use metadata necessary to enable the discovery, preservation, and reuse of their data (Tenopir et al., 2011).

Indian Council of Agricultural Research (2014) emphasizes that research data collection should be done scrupulously and data records should be maintained through proper use of metadata in a durable and accessible medium that ensures safety from tampering and manipulation. For example, crop improvement programs, to ensure traceability of the material and developmental process of the genetic materials for pedigree management, each line should be assigned unique ID in a season. All the data files should be named so that they clearly identify the data collected, research project, and so on. Proper documentation is needed to ensure that other researchers can use the datasets even after decades. One is able to decipher and use it as if (s) he had generated the data, and it is possible to prevent misuse, misinterpretation and confusion.

6.4.4 Storage, backups and preservation of research data

Storage, backups and preservation of research data are part of the stages in data curation and play a vital role in RDM. The section is divided into three sub-sections from 6.3.4.1 to 6.3.4.3.

6.4.4.1 Storage of research data

Data storage is important for making research data accessible to researchers in the future who wish to evaluate or augment the results of research carried out earlier. The accessibility of any data depends on the quality of the storage medium and the availability of the relevant data-reading equipment for that particular medium. DCC lifecycle model (Higgins, 2008) states that research data should be stored in a secure manner adhering to relevant standards.

The respondents were asked to state where they stored their research data in short term, after acquisitions. The findings indicated that majority 28(22.6%), 11(8.8%0), 21(17.0%) of the respondents from research institutes A, B and C respectively stored the research data on their laptop computers. Furthermore, in research institutes A, B and C, 24(19.3%), 13(10.5%) and 13(10.5%), of respondents respectively stated that they stored their research data on their research group's data storage file and also CD or DVDs. Some of the issues emerging from these findings related specifically to inadequate enforcement of RDM policy regarding standardized storage, absence of a coordinating unit on RDM, and absence of advocacy programs concerning standardized storage and its relevancy in data curation.

Further investigation on the storage media of research data preservation led to the realization that the majority of respondents preferred personal computers (107(86.3%) to store their data while 98(79%) preferred a hard drive of the instrument. Furthermore, 94.5% agreed that they stored their research data in a departmental server and institute's repository. These findings concur with the findings of the interviews where most respondents confirmed that they preferred backing up their data on their lap tops, external hard drives or in cloud services, research institutes server or repository and the reason for this was to maintain privacy and avoid plagiarism of their research data. Fary and Owen (2013) assert that as data moves through the various stages of its lifecycle, it is maintained in storage environments that provide varying levels of accessibility to the research community who may be interested in accessing data. At first, the data is contained in a private state, with only the research team having access. Then, the data is moved to a state that allow it to be shared inter-or intra-institutionally. Finally, the data is moved to an archive either directly from the private or from the shared environment. Fary and Owen mention that central IT storage, departmental storage environment, and cloud-based environment are some of the examples of storage environments that could house the data. A survey carried out by Mossink, Bijsterbosch and Nortier (2013) on Support Infrastructure Models for RDM (SIM4RDM) in Europe using positivist paradigm revealed that institutional repositories were being deployed to store finished datasets, for example, the Netherlands, the UK and Finland have well-established instructional repositories for research data.

6.4.4.2 Backups of research data

Backups of research data is essential for RDM to protect against accidental or malicious data loss due to hardware failure, software or media faults, virus infection or malicious hacking, power failure and human errors. University of South Africa Library (2017) asserts that to safeguard research data, both storage and backup are essential for three reasons: one, a granting agency may require retaining data for a given period and this should be explained well in a data management plan how data will be stored and backed up; two, storing and backing up data ensures that it will be there when it is needed for publications or grant proposal; and lastly, good storage and backing up help make data available to researchers in the research group, department, discipline or research institute in the future.

This study sought to determine the person responsible for the day-to-day management, storage and backup of the data arising from the research. The findings showed that there were many parties involved, for example the researchers, the research group's data manager, department, IT staff, librarian and the research unit. Furthermore the findings revealed that the majority of respondents in all research institutes preferred researchers, the research group's data manager and department to be responsible for the management, storage and backup of research data. The finding were supported by 22(17.8%), 12(9.6%) and 23(18.5%) of the respondents from research institute A, E and F respectively who stated that they preferred the researcher and another 25(20.1%), 13(10.5%) and 21(17.0%) of respondents from the same research institutes preferred the research group's data manager. Also the department was equally considered to be the most preferred to take charge in the day-to-day management, storage and backup of the data by all research institutes. The findings would appear to suggest that researchers were protective of their research data. However, findings seem to suggest that researchers were not aware of RDM policy or were merely ignoring it by becoming selfish of their data. Van Tuyl and Michalek (2015) in a study on assessing research data management practices of faculty at Carnegie Mellon University using qualitative epistemology found that faculty members backed up their research data on local computers, external hard drives, department of IT units or in cloud service. The same study also established that graduate and postdoctoral students were the individuals were actually responsible for day to day decisions around RDM.

The findings revealed that backups for the research data were done daily, weekly, and monthly in the institutes depending on individual researchers and research institutes policy. Majority of the respondents 89(71.8%) and 85(68.6%) preferred backing up their data on daily and weekly respectively and another 71(57.2%) made backups on hourly bases. However, there were other respondents who stated that they had never made backups. The finding are consistent with National University of Ireland Galway (n.d) that whenever a researcher makes changes to files, or adds new files there is a need to backup data and it is also a good practice to backup files on a daily basis and more often when working on critical research projects. This is because research data are vulnerable to loss when researchers upgrade their computers or software (Jahnke and Asher, 2012).

A study by Stanford Libraries (n.d(a)) established that creating multiple backup copies of your research data is an important part of data management, but it is far less effective when all those copies are stored in the same place and not well protected. Managing large files was found to present significant challenges for researchers in the research institutes because the infrastructure did not provide adequate storage space or sufficient bandwidth for data access. Agricultural research institutes should provide adequate technical infrastructure, for example, institute servers in different locations, repositories and secure cloud services to facilitate storage and back up of research data.

6.4.4.3 Preservation of research data

Ensuring access to and use of enduring data assets is a shared responsibility across agricultural research institutes. DCC lifecycle model (Higgins, 2008) underpins preservation as one of the stages in data curation which should ensure that data remains authentic, reliable and usable while maintaining its integrity. In this respect, an ideal digital preservation environment contains a mix of policies, processes and resources including staff and technologies to create an enabling environment for access, reuse and sharing of research data. Van den Eynden (2013) posits that data preservation entails migrating data to the best format and suitable medium, creating metadata and documentation for easy discoverability, backing up and storing, and finally archiving research data.

The findings revealed that the majority of respondents kept the research data for a period of more than ten years 25(20.2%) and 5-10 years 25(20.2%) respectively. In addition, the majority of respondents 29(23.4%) did not know the length of time research data was kept in the institute. This finding may suggest absence of an elaborate preservation plan and lack of awareness by researchers on what RDM preservation policies stipulate. These findings support the idea of University of Pretoria (2007) policy for the preservation and retention of research data which states that research data belongs to the research institution and must be stored for a minimum period of ten years after the completion of the research project but if intellectual property or contractual requirements are involved then the reservation period may be altered. A contrast is drawn from this finding with that of a study by Shakeri (2013) which revealed that 60.3% of the respondents indicated that research data should be preserved only for a maximum of five years after the results have been published and then disposed of permanently whereas 28.5% of the respondents indicated that they would preserve their research data for more than ten years because they needed it for reference and re-analysis.

The findings further revealed that researchers valued preserving data as manifested in the responses in which the majority of them in all the research institutes agreed that individual researchers, researchers in consultation with their researcher supervisor, and funder's requirements were the determinants of research data worth preserving. Twenty three (18.8%), 20(16.1% and 15(12.1%) respondents from research institutes A, C and E respectively agreed that research data preservation was decided by the researchers. Researchers in consultation with their researcher supervisor were also determinants of research data that were worth preserving with respondents 25(20.2%) and 20(16.2%) from research institutes A and F respectively. Furthermore, funder's requirements determined the worth of research data preservation in almost all the research institutes. These finding were not surprising given the absence of a coordinating unit which is the primary mandate and has statutory responsibility to manage research data. In addition, the findings would appear to show that enforcement of the preservation policy was not harmonized in the institutes. According to the University of Sheffield (2017), decision about preserving data should be done taking into account institutional RDM policy, funder requirements and data repository requirements. Furthermore, Higgins (2012) posits that activities that support the preservation process should be planned in a structured and organized manner through the maintenance of a preservation plan, which is to be reviewed and revised at regular intervals.

The findings also revealed that the majority of respondents preferred their research data to be moved to a secure archive for long-term preservations immediately after publication of a paper in a journal. This was according to 93(75%) of the respondents. Moreover, 88(71%) of the respondents stated that the time for moving research data to a secure archive for long-term preservations was determined by RDM policy. The findings of the study seem to suggest that preservation policy, where it existed was either ignored by researchers or was not enforced. Tenopir, Birch and Allard (2012) opined that research institutes need to provide not only structure and policies for research data preservation, but services to support and educate researchers on concepts of data management and promote preservation of data sets that is vital for the continuation of research. The preservation policy should provide framework for guiding researchers on preserving data.

The findings established that respondents had various ways of archiving research data for long-term preservation. Majority of the respondents in various research institutes archived research data as supplementary files attached to the journal articles on publisher's website as indicated by 27(21.8%) and 17(13.7%) respondents from research institutes A and C respectively. Other respondents preferred archiving their research data in the research institute's data archive server or databank repository. The findings seem to suggest that the researchers were not compelled by any policy to archive data in a particular preservation place. The implication of this finding is the possibility that it could accelerate loss, inadequate access and sharing of data due to the absence of systematic archiving of research data. Witt (2008) posits that datasets should be preserved in institutional data repositories, funder established data centre or domain-specific repositories which could enable long-term access, its discovery and use in the future. In the same vein, National Institutes of Health (NIH) (2008) and National Science Foundation (NSF) (2007b) are beginning to require the deposit of the publication derived from research they have sponsored into repositories. Similarly, they are also requiring that grant proposals include data management plans that address preservation and open access to data that is generated by projects they have sponsored. Preservation strategies are becoming important instruments in every organization or research institutes to enhance access, reuse and sharing of research data. Shakeri (2013) noted that publishers were preserving research data associated with their articles as supplementary material because they had not employed any data preservation strategies other than making back-up copies of the data. Access to data preserved by the publisher in agricultural research environment could be challenging especially to researchers who access, reuse and share research data from time to time. Therefore, preservation and archiving of data for management, access, reuse and sharing in the distant future necessitates preservation strategies, human capability, economical, and technological sustainable systems.

Moreover, the findings revealed that researchers preferred transferring research data on physical hard drives and by web-based selected data repositories. Transferring data through e-mailing file to the librarians and IT department was not considered as much because most researchers did not value the role of the two departments in RDM. The findings seem to suggest that librarians and IT department were relegated in RDM. Jahnke and Asher (2013) explain that although librarians and IT staff may not take on full responsibility in RDM, they can introduce services to support and develop datasets for easy accessibility and sharing.

The study also sought to understand the responsibility for research data that the researchers leave behind when they exit the research institute. The findings revealed that the majority of respondents from all research institutes seemed to leave the research data with their supervisor or research institute. According to 19(15.3%) and 18(14.6%) of the respondents from research institutes C and F respectively, preferred leaving the data with their supervisor. Leaving the research data with the research institute was expected because RDM and intellectual property policy where they existed were clear that any data generated from research project sponsored by the research institute or funded through the institute, belongs to the institutes (University of Pretoria, 2007). RECODE Project Consortium (2014) emphasizes that policies and regulations regarding retention periods and where to deposit research data when researchers leave the research institute, should clearly be stipulated. Digital preservation and archiving of research data ensures that management of digital research data is facilitated over time with a view to retaining their intellectual content, authenticity, and accessibility for a variety of uses (Western Libraries, 2012).

6.4.5 Research data access, sharing and reuse

The value of research data lies in their use. Organization for Economic Co-operation and Development (OECD) (2007) outlines the benefits of improved access to, sharing and reusing of data namely: reinforcing open scientific inquiry; encouraging diversity of analysis and opinion; promoting new research; enabling the exploration of topics not envisioned by the initial investigators; permitting the creation of new data sets when data from multiple sources are combined; and providing greater returns from the public investment in research.

6.4.5.1 Research data access

Open access to research data refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. Depositing research data in repository and providing open access to publications are two main routes to open access to research data (European Commission, 2016; OECD, 2007, 2004; and Berlin Declaration, 2003). A DCC lifecycle model (Higgins, 2008) point out that access to research data ensures that data is accessible to both designated users and reusers. Similarly, CCM framework (Lyon et al, 2012) emphasizes that for data to be accessible, data formats, collection methods, description, and data packaging and transfer protocols should be considered.

The respondents were asked to state the time taken to halt their research data before it is published. Majority of the respondents 84(67.7%) stated that they wait until the journal article describing the research results has been published whereas other researchers 67(54%) will give a period of one year, to permit them to exploit their research results. Thanos (2010) asserts that, researchers avoid premature publication in order to develop a strong argument before going public and also protect information that should be treated as confidential. Open access to agricultural research data from public funding should be easy, and user-friendly.

The findings further revealed that most researchers from agricultural research institutes restricted public access to their research data because of various reasons including: avoid prior disclosure, confidentiality of research data as per the agreement between the research group and the commercial partner sponsoring research, research data being confidential, proprietary or classified, intellectual property concerns, or lack of appropriate tools for accessing or publishing data.

Furthermore findings indicated that the majority of people who were currently accessing research data were group researchers with 56(45.2%) respondents agreeing. However, 37(29.8%) and 21(16.9%) respondents revealed that other researchers in the research institute and researchers from other research institute respectively were currently accessing the research data. The findings also revealed that there was minimal access to research data by the public. The findings seemed to suggest that there were inadequate mechanisms to facilitate access to data. In support of the findings, European Commission (2016) posits that fuller and wider access to research data is important as it helps to build on previous research results, encourage collaborations and help avoid duplication of effort, speed up innovation, and involve citizens and society in the scientific process.

Access to Kenya's agricultural research data faces various challenges as enumerated by Muinde and Gorman (2009) that include: social-cultural (non-visionary leadership), lack of ICT infrastructure, legal, policy and institutional frameworks, and capacity building programs. Muinde and Gorman (2009); and Chisenga (2012) point out that there is a necessity for government to fund, plan and prioritize resource allocation for research to ensure that there is local content which is visible, accessible and sharable online to facilitate information flow.

The study also established that the majority of researchers in Kenya's agricultural research institutes availed their research data through published literature and this was attested to by many journal publication in various research institutes sampled for the study, for example the Tea and Coffee journal, among others. Conferences, seminars and workshops were considered as another method of availing research data for open access and this was regarded as the best method of accessing and sharing data and information. Institutional repositories and media were averagely used because the researchers were skeptical about privacy and plagiarism of their research data. These findings are supported by Henty (2014) who concurs that open access is encouraging researchers to make their publication available in institutional repositories. A study by Tenopir, Birch and Allard (2012) using qualitative and quantitative epistemologies revealed that researchers avail research data to others through institutional repositories, conferences, seminars and workshops; collaborative web space, data portal or database drive website, external storage device, hard copy or print. Kedemi (2017) noted that the implementation of KAINet, an institutional/national repository with scientific publications on agriculture and forestry, faced a number of challenges which included: absence of institutional policies that support open access, the low awareness of copyright issues, absences of appropriative information, management skills, system incompatibility, shortage of technical ICT skills and collaborations with stakeholders.

A study done by Jahnke and Asher (2012) revealed that improved privacy and data access control were needed because they are essential to develop tools that manage confidential data and provide the necessary security. Most importantly, access policies must be developed and enforced to ensure that researchers have control over data, as well as over who has access to it. Without such assurance, many researchers are unlikely to invest in open access systems. The call for open access to research data is in tandem with the Berlin Declaration on Open Access (2003), which states, that results of research produced must be made widely available on the internet for users to use and reuse in order to accelerate the pace of scholarship and research.

6.4.5.2 Sharing research data

The sharing of research data has long been practiced among many research communities. This has been made increasingly easy with the advent of internet tools such as emails, web portal, web sites, data repositories, groupware among others. CCM framework (Lyon et al., 2012) pointed out that data formats; processing workflows; packaging and transfer protocols; data description; vocabularies, semantics and ontology; and data identifiers are key in data sharing.

The present study established that the majority of researchers shared their research data with members of their research group and trusted external collaborators, their research supervisors, and research sponsors before publication of any papers arising from research data interpretation. This was supported by 26(21.0%) and 17(13.7%) respondents from research institutes A and C respectively who agreed that they shared their research data with members of their research group and trusted external collaborators, and 12(9.7%) and 19(15.4%) respondents from research institute E and F respectively who agreed that they shared research data with their supervisor. Further, 21(17.0%) and 12(9.7%) of the respondents from research institutes A and B respectively shared research results with research sponsors. Some of the issues that emerged from these findings included limited mechanisms for sharing data especially with the general public. Warren (2016) appreciates that there are many policy, privacy and practical issues that need to be addressed in order to make sharing of research results practical and useful in the research institutes. Tenopir, Birch and Allarch (2012) are of the opinion that the decision of researchers to share or not to share data is quite personal due to factors such as privacy concerns, concerns about publishing opportunities, and the desire to retain exclusive rights to data.

The findings showed that researchers in agricultural research institutes shared research data using different methods and the most preferred methods were: use of external storage device, hard copy/print, e-mail, institutional repository, and journal publication. From the results in Table 5.21, collaborative web space, data portal/database driven web site, and depositing them with a specialized data centre were methods that were partially used in sharing research data. The findings are in agreement with Van den Eynden et al., (2011) who found that most research results were shared through specialist data centers, data archives or data banks; journal publications; institutional repository; institutional websites; and informal sharing between researchers on a peer-to peer basis. Van den Eynden et al maintain that sharing policies, embracing new sharing technologies and training researchers are key issues in sharing research data.

The study sought further to investigate the benefits of sharing research data and found that majority 110(88.7%) of the respondents benefited from sharing research data because it encouraged scientific enquiry and debates, and also reduced the cost of duplicating data collection. In addition, 108(87.1%), 106(85.1%), and 104(83.8%) of the respondents benefited from sharing data through increasing the impact and visibility of research, enabling scrutiny of research findings, and leading to new collaborations between data users and data creators respectively. This finding resonates with the finding of a study done by Borgman (2012) which revealed four reasons for sharing research data: reproduce or verify research; make results of publicly funded research available to the public; enable others to ask new questions of extant data; and advance the state of research and innovation. Steinhart, Chen, Arguillas, Dietrich and Kramer (2012) also assert that sharing agricultural research data has the potential to facilitate collaborative approaches in conducting research and when data is shared more widely, it holds the potential to advance knowledge within a given discipline and even across disciplines.

As much as there are benefits there are also challenges in data sharing. The respondents who participated in this study stated that the legal issues relating to patent privacy was a major challenge in data sharing. Further, 100(80.6%) of the respondents stated that there were fears regarding misuse of shared data and building trust. Enke et al., (2012) identified the factors that could impede agricultural researchers from sharing data to include: fear of loss of control over the data; lack of research institute's-wide standards for data sharing; the amount of time that would be needed to invest in sharing data sets; researchers concerns with legal issues; misuse of data and incompatible data types; and researchers' lack of knowledge required for constructing an articulate data management plan that could assist in sharing data with the funding agencies on the same platform. This is supported by Karasti, Baker and Hakola (2006) who point out that there is a need for agricultural research institutes to provide not only structures and policies for research data sharing, but services to support and educate researchers on concepts of data management and strategies for sharing data that is vital for RDM. Widespread sharing of data may lead to discovery and use outside the discipline in which the data was created, fostering interdisciplinary research and learning.

6.4.5.3 Reusing research data

Data created from research are valuable resource that can be used and reused for future scientific and educational purposes. DCC lifecycle model (Higgins, 2008) encourages access, use and reuse of data by ensuring that data is accessible to both designated users and reusers, on a day-to- day-day basis.

Concerning the use and reuse of research data, the majority of respondents stated that research data is used and reused in order to promote potential new data uses and encourage scientific inquiry. In addition re-analysis of data leads to powerful insights, promotes innovations, and helps avoid duplication. These findings corroborate those of Australian National Data Services (n.d (a) in a study on data reuse which reiterated the reasons for enabling reuse of data to be: encouraging scientific enquiry and debate, increasing the impact and visibility of research, providing great resources for education and training, and leading to new collaborations between data users and data creators. Lewis (2010) agrees that the reuse of research data assists in addressing emerging issues, supporting re-analysis of existing data or comparisons with new data in order to come up with new research themes and powerful insights which are imperative to RDM in agricultural research institutes.

From the perspective of using and reusing research data, majority of the respondents agreed that research data is made available for use and reuse through publications as evidenced by 32(25.8%) and 19(15.4%) respondents from research institutes A and C respectively. Another 15(12.1%) and 18(14.5%) respondents from research institutes B and F correspondently agreed that research data is made available through citations. Other respondents agreed that their data was made available through sufficient metadata that described how the data had been specified, collected, analyzed and transformed. Van Wyk and Van der Walt (2014) noted that research data that is made available either through publication or citations can be used and reused to follow-up research, rediscover new research areas, undertake research reviews, scrutinize findings, and be used for teaching and learning.

Responsibility for RDM in agricultural research institute is another aspect that was the focus of the study. The findings indicated that there was no clear unit/department/person responsible for RDM. For example, in research institute A different respondents 13(10.5%), 13(10.5%), 14(11.3%) and 15(12.1%) agreed that IT staff, librarian, collaborative responsibility/research group, and external research parents respectively were responsible for RDM in agricultural research institutes. The finding seem to suggest that in Kenya's agricultural research institutes, there was no clear policy regarding responsibility of RDM and this could explain why research data remained underdeveloped, underutilized and with no budgetary and resource allocation. Tenopir, Sandusky, Allard, and Birch (2014) acknowledge that there has been an increasing need for libraries and librarian consultation with IT staff to play a leading role in RDM. Lewis (2010; Erway (2013); and Tenopir et al., (2014) further notes that the library is well situated to be a key player and provide a range of service in RDM that include data curation, access, sharing of data literacy to researchers, institutional repositories, among others, given its extensive experience.

6.5 RDM knowledge, skills and training requirements

RDM knowledge, skills and training requirements needed to capture, appraise, describe, preserve, access and reuse research data that was studied using the CCM framework and the DCC lifecycle model as the analytical lens. RDM knowledge, skills and training are catalysts for facilitating the success of research data curation. The DCC lifecycle model (Higgins, 2008) is used to plan activities and map out roles and responsibilities in data curation. The CCM framework (Lyon et al., 2012) on the other hand, provides human resource capabilities to ensure that all necessary activities, roles and responsibilities in all stages of data curation are undertaken. The discussion on these aspects is provided in sections 6.4.1 - 6.4.5 respectively.

6.5.1 Knowledge and skills available RDM

The findings established that the majority of researchers who participated in the study had knowledge of open access 94(75.85), research data sharing 91(73.3%), preservation 86(69.35), and data curation 82(66.1%). Most of the researchers were knowledgeable in relation to open access and data sharing because they were engaged daily in accessing and sharing data to enhance their research with their research groups, supervisors or collaborators. The majority of researchers did not posses knowledge on data curation and preservation. A possible explanation for these findings could be that researchers were not aware of the data curation lifecycle which is critical to RDM. Schmidt and Shearer (2016), in a study of librarians' competencies profile for research data management, guided by interpretive paradigm, enumerated the knowledge that RDM staff should posses must include: knowledge of repositories, data manipulation, data discovery mechanisms, funders' policies and requirements, data centers, data publication requirements of journals, sharing and access, data citation and referencing, metadata standard and schemas among others. In addition, Gold (2007) claims that it makes much more sense to train domain experts in curation skills than it does to try to teach non-domain librarians or archivist to understand the infrastructure and service needs of a domain.

6.5.2 Skills available for RDM

RDM skill is another aspect that is vital for RDM. The findings of the study revealed that researchers were more skilled in collaboration and communication, sharing, tools and technologies, preservation, data capturing, searching and retrieval; however researchers had little skills on metadata. The findings seemed to suggest that due to inadequate skill in data curation lifecycle stages, RDM was not executed effectively in the research institutes. Fary and Owen (2013); and Creamer, Morales and Crespo (2012), in their studies on RDM skills and competencies listed the skills relevant to RDM which included: storage, data migration, networking, legal, financial, security, metadata creation and assignment, scholarly data communications, and preservation. Kennan (2016) interviewed 25 data professionals in Australian scientific research organization and found out that the most common set of skills required were: interpersonal skills, data specific knowledge and skills, and metadata. In this regard, it would seem knowledge and skills possessed by researchers and other RDM stakeholders in Kenya's agricultural research institutes were inadequate.

6.5.2.1 Skills gaps in RDM

The study found that the majority of researchers had skills gaps in the use of institutional repository 103(83%), while 101(81.4%) of respondents had gaps in the use of ICTs tools and equipment, in data curation and in complying with the various mandates of funders. Developing metadata scheme had fewer respondents with the skills gap. From the results, it appears that all the respondents from the agricultural research institutes surveyed had similar skills gaps. This result could be attributed to the absence of RDM orientation, user education, and advocacy campaigns in Kenya's agricultural research institute. Lyon (2012); and Lewis (2010) noted that innovative approaches were needed to address the significant skills gaps, data literacy and training in areas of data curation, open access, sharing, research data services, use of institutional repositories, metadata, use of ICT, and reuse of research data. In addition, Kahn et al., (2014) concurred that skills, knowledge and resource gaps in the agriculture research institutes needed to be addressed using training providers who have the expertise and capacity.

6.5.3 Training needs

The respondents noted that training was needed by researchers, on ICTs, metadata management, security and storage of data, data curation (data capture, appraisal, description, preservation, access and reuse), collaborative partnerships, funder requirements for research data, data management plans, data sharing as well as legal and ethical consideration in RDM. These findings corroborate the findings of the interview where heads of research, librarians, and IT specialist confirmed that the researchers and the heads of departments needed training in library and IT skills which enable them appraise, describe, preserve, search and retrieve, access, share, and use RDM system, among others. Jahnke and Asher (2012) in a study on the problems of data management among university researchers revealed that the majority of researchers had not received formal training in data management practices, and were not satisfied with their level of expertise though they acknowledged they were learning on the job albeit in an ad hoc fashion. Another study by Mukiibi (2016) acknowledged that most Ugandan universities offering Computing, Information Science and other related programs to RDM lacked RDM studies in their academic programs thus inhibiting the development of human capacity to do RDM. Besides, Lewis (2010); and Flores et al., (2015) recommended that the head of RDM unit/department in identifying the skills gap should work in partnership with library and information science (LIS) schools to develop new training and development resources to fill it. Lötter, 2014); and Patrick et al (2013) pointed out that RDM training should be undertake through capacity building workshops; in-house training and mentorship of research data curators. In addition, opportunities should be made available for researchers to attend conferences on research data and information access, sharing including data curation and networking.

6.5.4 Satisfaction with quality of research data services

The study found that majority of respondents were satisfied with the quality of RDM services provided by librarian and IT specialists in such areas as reference and advisory support for researcher on RDM; technical support for research data service systems (repository, web portal, creating metadata, RDM system, access and discovery systems). On the other hand, other respondents were dissatisfied with research data services offered by the library and IT departments in the following areas: creating web guides and finding aids for data repositories, provision of RDM literacy programs, and coordinating RDM advocacy programs between the researchers and the stakeholders. These findings substantiate the finding of the interview with librarians and IT staff, where most respondents confirmed that Kenya's agricultural research institutes have libraries that were rarely used by researchers because: they offered few services related to RDM, the information resources available were outdated, absence of knowledgeable and skilled librarians in RDM, poorly equipped with ICT infrastructure, absence of policy mandating the library to support RDM. Equally, IT departments had a very small role to play in RDM and in most cases assisted with trouble shooting, repairing researcher's faulty computers, carried out maintenance of web page and channels of communication (email, internet).

Barber and Zauha (1995) underscored the role librarians can play in RDM because they have expertise in different areas like the classification and description of information, metadata services, cataloguing and indexing, information literacy, selection, publishing and scholarly communication, and information formatting (repackaging). In addition, librarians have been on the forefront in adopting new and electronic information formats, library management systems, preserving information, and providing access to information. Therefore libraries with the assistance of IT department can help agricultural research institutes with research data curation, maintaining collections, organizing data literacy programs, advocacy and outreach (Searle, 2011). Meanwhile, few researchers are aware of the data services that the library might be able to provide and seem to regard the library as a dispensary of books and articles rather than as a locus for real-time research/professional support (Jahnke and Asher, 2012).

6.5.5 Data literacy programs

The findings showed that the majority of researchers had not undertaken any research data literacy program in the research institutes surveyed as revealed by 95(76.6%) respondents. This finding corroborates the findings of the interview where most respondents confirmed that researchers, heads of research, librarians, IT staff and other RDM stakeholder had not undertaken any data literacy programs. The findings seem to suggest that RDM policy inadequately addresses data literacy strategies in the research institutes. Peters and Dryden (2011), in a survey using DCC lifecycle model, found that the top data services that researchers needed were primarily directional ones like: assistance with data management plans and the proposal process, data curation, sharing, reusing, finding data-related service, publication support, and targeted research assistance with data management.

6.6 Level of ICT preparedness in RDM

The growing capacity of ICTs has contributed immensely to research especially in inventing new ways of conducting research. ICT incorporates electronic technologies and techniques used to manage research data, including data and information handling tools used to capture, describe, preserve, process, distribute and exchange information. ICT preparedness in Kenya's agricultural research institutes was studied using DCC lifecycle model and CCM framework as the analytical lens. DCC lifecycle model is used to build frameworks of standards and technologies in every stage of the lifecycle while CCM framework focuses more on the types and requirements of ICT tools and equipment needed in RDM. The discussion on these aspects is provided in section .6.5.1 to 6.5.3 respectively.

6.6.1 ICT tools and equipment

ICT tools and equipment is another aspect that is imperative in RDM. Henty (2014) stated that the growing contribution of ICT(s) to research has excited researchers the world over as they invest in new ways of conducting research while enjoying the benefits of more sophisticated computers and communications systems that support measurement, analysis, modeling, simulation, collaboration and publishing.

The ICT tools and equipment used in RDM was another issue of concern in the study. The respondents were asked to name the ICT tools and equipment used in research data capture, appraisal, description, preservation, access, and sharing. The study established that researchers did not have adequate ICT tools and equipment to facilitate RDM despite the fact that KALRO's strategic objectives include *"enhancing availability of knowledge, information, and technologies on agricultural and livestock product value chain".* The study revealed that ICT tools and equipment majorly used in data curation in the research institutes were laptop computers, audio recorders, GIS, Video, cameras, scanners, institutional repositories, servers, flash disks, hard disk, emails, publications, websites and others. The findings further revealed that in areas of appraisal and description, ICT tools and equipment had not been embraced and for this reason researchers did not have a standardized way of describing research data.

These findings concur with the interview where most respondents confirmed that computers, especially laptop computers were very common among researchers and technicians in the research institutes as they were used to capture, analyze, preserve, access and share research data. In some institutes for example research institute B, C and F respectively, tablets and GPS were used to relay research data in real time. It was also revealed that scanners, cameras and mobile apps were used in all research institutes. Furthermore, the results showed that data analysis packages such as SPSS and Statistical Analysis System (SAS) were used to analyze data captured in the field.

The finding of a study by Jahnke and Asher (2012) on problems of data management and curation practices among university researchers showed that researchers had encountered challenges with a variety of technical issues concerning RDM. Such challenges were: inadequate access to networked storage, data loss because of poor organization, file format, and the scale of their data overwhelmed available infrastructure. Witt (2008) posits that while cyberinfrastructure has been revolutionizing digital research, a comprehensive framework for capturing, organizing, preserving, and making research data available and usable has not been created. Access to adequate ICT tool and equipment should be in tandem with the training of researchers and other data professionals on how to utilize them effectively in RDM.

6.6.2 RDM software(s) used

The results on RDM software used in the research institutes revealed that different agricultural research institutes had different software for data analysis because of the different diverse research projects being done. Most of the software packages used in agricultural research institutes was statistical software such as SPSS, SAS and Genstart used to analyze research data. However, research institute E had acquired RDM software called Grain Global to facilitate RDM and provide research data services like web portab, metadata, web 2.0, repository, and intranet on plant, animal and microbial genetic resources research data. Studies done by Jetten (2014); Maru (2004); European Strategy Forum on Research Infrastructure (2009); and Fary and Owen (2013) found that government and agricultural research institutes are keen to support research in order to ensure that research data are well managed, readily accessible and available on open access. They recommend that if government is paying for the research, then the public should be entitled to have access to all the products of research facilitated through ICT. Consequently, the use of web 2.0 by

Kenya's agricultural research institutes should be supported to enhance visibility and exchange of research outputs, including metadata, and sharing research outputs (CIARD, 2012).

6.6.3 Frequently used data security methods

The respondents were asked to state the frequently used data security methods in their institutes. The findings show that majority of respondents use physical security 47(37.9%), followed by network security (41(33.1%), and lastly encryption 36(29%) for data security. A possible explanation for this might be that researchers preferred physical security for the reason that it was within reach and more trusted by the majority of researchers The findings correspond with the interviews administered to directors of institutes, heads of research, librarians and IT specialist who revealed that research institutes had safes to keep sensitive and classified information. In addition they used unique passwords, multiple hard drives and cloud for back up. However, some of the issues emerging from these findings related specifically to absence of security policy and inadequate awareness on the use of encryption method and other methods like cloud services and security software to protect research data. In relation to this finding, Amorim et al. (2015) emphasize that there is a need for compatibility of research data security with data repositories, metadata, security systems, data management systems, and search mechanisms in order to enhance privacy of research data. Moreover, data security is important for protecting intellectual property rights, commercial interests, or to keeping personal or sensitive information safe (Van den Eynden et al., 2011).

The level of ICT(s) infrastructure preparedness with regard to RDM was rated between 50% and 80%. This level of preparedness was generally acceptable and was attributed to availability of internet in all Kenya's agricultural research institutes, availability of laptop computers to all researchers, and access to institutional repository among others. Mugabe (2001) in a study affirms that ICTs are creating faster ways of acquiring, storing and disseminating information, thus breaking barriers to knowledge and facilitating integration into global economy. Nevertheless, lack of and inappropriate ICT tools and equipment has created weak linkages between researchers, extension workers and farmers resulting in a major constrain that has resulted in research findings not being applied by farmers (Munyua, 2000). This view is consistent with the argument advanced by Kassam and Odame (2002) that the full potential of agricultural research is not being realized because the government and research institutions have not embraced ICT tools and equipment in agriculture thus inhibiting communication between scientists, extension workers and farmers throughout developing countries.

Therefore, Maru (2004) in a study sums up gaps and weaknesses of ICT used in Kenya's agricultural research institutes to be as follows: Capacity (including infrastructure and skill); Content (including generation and management); Capital (with the focus on funding not only ICT also capacity building); connectivity (not only physical but the ability to access information by individual and user community); and collaboration (within and across research institutes at national, regional and international).

6.7 Collaborative partnerships in RDM

Collaboration within institution and among institutions is necessary for the management, access, reuse and sharing of research data as well as for creating and sustaining public-private partnerships among research institutes (Humphrey, 2014). Collaborative partnership in this study is underpinned by CCM framework (Lyon et al., 2012) by focusing on collaboration within the discipline, across disciplines, across sectors and public. Collaborative partnerships can also exist within research institutes, across institutes, nationally, regionally and internationally (Lyon et al., 2012). Collaborative partnerships play a key role in RDM because it increases the probability that the knowledge, skills, techniques, access and sharing of research data required will be available within the collaborators.

6.7.1 Collaborations within the discipline/sector

This study sought to establish how research data capture, appraisal, description, preservation, access and reuse in Kenya's agricultural research institute influenced collaborations within the discipline. The results revealed that majority of the respondents were satisfied with collaboration with departmental research groups, collaboration across research group between organizations, and the international collaborations. Nevertheless, the respondents were dissatisfied with collaborations organized at national level. The findings seemed to suggest that there was inadequate government support in terms of budgetary allocation to do research (generate research output to share at national level), organize and attend conferences, seminars and workshops at national level. The findings are supported by Muinde and Gorman (2009) in a study on barriers to open access to agricultural information in Kenya who found that government planning and resource allocation to agricultural research institutes was poor owing to lack of research funds which limited research and in turn resulted to inadequate research output. In this regard, collaborative partnerships become a challenge since there is no research data to access, share and enhance RDM.

6.7.2 Collaborations and interaction across discipline

The study also established that the respondents were satisfied with individual researchers occasionally collaborating outside their discipline, interactions across discipline collaborating through joint conferences or publications, and formal collaboration between research groups from different disciplines. The findings suggest that researchers were satisfied with collaborations across disciplines especially joint conferences and publication, seminars and workshops, among others. A good example is research institute B which had various publications like growers handbook 5th edition, annual technical reports, and journals published twice per year that was shared across disciplines in all research institutes which enhanced access and sharing of research data. Collaborations across discipline are vital since they promote new insights if the research data is combined and when the opportunities and advantages of these collaborations stem largely from changing expectations and environment for research (Lyon et al., 2012).

Concerning collaboration with the public, the findings also showed that the respondents were satisfied with informational or participative media programs organized to engage the public. This was in agreement with interviews done with the heads of research which also revealed that collaborative partnerships with the public through trade fairs, pamphlets, brochures, agricultural shows were highly accessed and shared. Collaboration between researchers and the public contributes significantly to increased productivity, quality of agricultural products, and diversified crops and livestock (KALRO- Tea Research Institute, 2016).

6.7.3 Partnership: geographical scale of funding for research

Concerning the level of satisfaction with the geographical scale of funding for research, the respondents were satisfied with research funded through grants from agencies and research funding by international bodies. Conversely, researchers were dissatisfied with research funded by government. This finding was largely attributed to inadequate support by government in as far as funding for research is concerned (Maru, 2004; and Muinde and Gorman, 2009). Meanwhile, Pinfield, Cox and Smith (2014) conducted interviews with 26 respondents regarding collaborative partnerships. The findings revealed that collaborative partnerships have benefits such as metadata exchange, exchange RDM human capacity, sharing and reuse of research data on one hand and on the other hand challenges such as lack of teamwork and policies governing collaborative partnerships.

6.7.4 Public-private partnerships

The majority of respondents were satisfied with informal partnership with public and private industry where no funding was involved, or research co-funded by public-private industry, and where partnerships involved formal co-investment partnerships running long-term multi-phase research. These findings validate the interviews where most respondents confirmed that agricultural research institutes were collaborating with universities; international and regional agricultural research institutes such as ILRI, ICRAF, Delamere dairies, and government ministries (especially ministry of agriculture and livestock) in research projects, data access and sharing. The findings appear to suggest that Kenya's agricultural research institutes were supported by public-private partnership in form of funding, capacity building, and technical infrastructure. (The World Agroforestry Centre, 2012; Chisenga, 2012; and CIARD, 2012). Jahnke and Asher (2012) notes that there is a great need for more effective collaboration tools, as well as online spaces that support the volume of data generated and provide appropriate privacy and access controls.

6.7.5 Benefits of collaborative partnerships

The respondents were asked to state the benefits that accrued from collaborative partnerships. Directors, heads of research, heads of IT and librarians interviewed, noted that through collaborative partnerships, they are able to access and share research data, share research tools and equipment, repositories, human expertise, statistical data, learn new skills, build and empower those without resources and ICT infrastructure. In addition, they indicated that joint publications, networking, joint innovation, and patenting are made possible through collaborative partnerships. The sentiments by the respondents interviewed were in agreement with a study done by Maru (2004) on agricultural research and development in sub-Saharan Africa which outlined the advantages of collaborative partnerships among agricultural research institutes at present and the near future to include:

- Increased and improved sharing and exchange of data, information, knowledge, skills, technology and resources, including financial, laboratory facilities, and experimental farms for multi-disciplinary and multi-location research;

- Increased donor, government and private sector funding for improving, further enabling and enhancing agricultural research systems at the national, regional and international level;
- Increased capacity to influence national, regional and international policies and strategies related to agricultural research, information systems and development.

Underscoring the importance of collaborative partnerships, Kahn et al. (2014) assert that collaborative research by international agencies and universities in the west have enriched research data repositories in their respective countries as per requirements of their funding agencies. Such collaborative partnerships could help Kenya's agricultural research institutes to facilitate long term preservation, access, reusing and sharing of research data resulting into new meaning and subsequently new findings which gives value to RDM (Mukiibi, 2016).

6.7.6 Data management plan

The findings established that researchers did have a DMP in the form of Standard Operational Procedures (SOPs). Most SOPs developed are comprehensive because the guidelines are provided for by ISO 9001:2005. However, SOPs did not include issues such as preservation, access and sharing which underpin RDM. DMP is underpinned by DCC lifecycle model (Higgins, 2008) and focuses on how research data are to be handled during a research project and after the project. Boston University (n.d) affirms that DMP is essential in RDM because it: ensures that data are in the correct format, organized well and better annotated; increases research efficiency; facilitates data archiving and preservation; ensures that research data and records are accurate, complete, authentic and reliable; ensures research integrity and replication; and enhances data security and minimize the risk of data loss.

6.8 Benefits, challenges and recommendations for RDM

The subsequent sections address benefits, challenges and recommendations of RDM.

6.8.1 Benefits of RDM

RDM is regarded as an essential enabler in the knowledge based economy. The findings of the study revealed that Kenya's agricultural research institutes were not optimally benefiting from RDM. The findings revealed the following benefits of RDM: properly practiced data curation would facilitate management, access, reusing and sharing as Dora and Kumar (2015) assert that, by opening up research data it will enhance visibility and attract new collaborators and research partners nationally and internationally. RDM would bring great opportunities of improving the pace and effectiveness of scholarly inquiry especially when relevant research data is discovered, reused and recombined (Lynch, 2014). RDM would eliminate duplication of doing research hence reduce cost of doing research through collaboration between institutions, groups and individuals (Kahn et al., 2014). Further, other respondents pointed out that RDM leads to re-analysis of data which in turn results in new innovations and accelerates the generation of more new knowledge. In the same vein, availability of RDM legal framework, policies and regulations provide best practices in RDM and also enhance security of data. Liaison between librarians and IT department in providing technical expertise on RDM would be made mandatory to enhance and improve standardization of RDM in agricultural research institutes. In addition, the University of Sheffield (n.d) states that sound RDM increases research exposure and impact; improves quantity and quality of research output; and increases collaboration, among others.

6.8.2 Challenges encountered in RDM

The benefits of RDM outlined above do not come without challenges, the finding revealed that Kenya's agricultural research institutes are still struggling with RDM challenges such as: absences of a coherent RDM legal framework and outdated policies and regulations which hampers the functions and activities of RDM hence affecting the completeness and quality of research data in agricultural research institutes (Jeffery, 2012; and Brown, Bruce and Kernohan, 2015). There exists an absence of qualified staff with RDM knowledge and skill to facilitate data curation and offer RDM services, especially where researchers are reluctant to engage in RDM (Jeffery, 2012; and Just & Whitaker). Inadequate data literacy and advocacy campaigns about RDM aggravate user education and orientation on researchers and other RDM stakeholders. The generation of large volumes of research data on a daily basis in agricultural research institutes have inadequate technical infrastructure for data curation for storing, processing and disseminating research data which exacerbates access and sharing of research data (Njuguna and Itegi, 2013). Further, financial constraint is another challenge that impedes RDM as Muinde and Gorman (2009); and Njuguna and Itegi (2013) assert that inadequate funding impact virtually all aspects of research including its mission, processes, participants' integrity and dissemination of findings. Moreover, absence of collaborative research data access and sharing inhibits visibility, access and sharing of research data which results in minimal collaborations among researchers. Absence of government and agricultural research institutes support to initiate RDM units encumbers the implementation and growth of RDM. Erway (2013) opined that limited involvement of librarians and IT departments in offering technical support in RDM denies maximum utilization of research data. Leigh (2014) and Sherbak (2014) in study found that common RDM challenges faced by agricultural research institutes included defining research data; logistics of managing research data; absence of institutional engagement; inadequate infrastructure; inadequate knowledge and skill on data curation; absence of capacity; and need for service development. Similarly, Cox, Verbaan and Sen (2012a) outline the RDM challenges to include: absence of institutional RDM policy; inadequate advisory services on RDM to raise awareness; inadequate data literacy to researchers and other related data professional; and absence on guidance regarding copyright and licensing of data as well as intellectual property rights.

6.8.3 Recommendations to mitigate the challenges

From the foregoing, Kenya's agricultural research institutes are facing several challenges in RDM. The respondents proposed some interventions to improve RDM in Kenya's research institutes that include: need for government support through enactment of RDM legal framework and funding where the agricultural research institutes can diversify ways of funding by intensifying research and producing agricultural research data which is usable to society and able to attract the attention of funders who will be willing to invest in RDM (Njuguna and Itegi, 2013) rather than relying on government funding. Appropriate ICT infrastructure that provide robust but cost-effective tools and equipment that are capable of processing, storing, accessing and sharing agricultural research data (Biddick, 2012). Recruitment of human resources with knowledge and skills in RDM as well training the existing staff is required. In addition, reviewing and reformulating RDM policies and regulation which are clear, comprehensive and coherent should be prioritized in order to enhance RDM (Hodson and Jones, 2013). Organizing data literacy and advocacy programs can be unraveled by sensitizing and incentivizing researchers and RDM stakeholders (Jeffery, 2012). RDM unit to oversee activities and functions of RDM in every research institute should be established. Enhance collaborative partnership in RDM to allow access and sharing of research data through RDM across-institutional collaborations where different teams work together to achieve a coordinated approach (Brown, Bruce and Kernohan, 2015). Lötter (2014) outlines the following measures to bring improvement in RDM: lobby for national policies and support for RDM; continued advocacy about RDM in organizations; mutual assistance in capacity building, sharing experiences, knowledge, support shared goals; encourage each other by celebrating achievements.

6.9 Limitations of the findings

The findings contained limitations that the researcher had to contend with given that RDM in agricultural research institutes is a relatively new study (Kahn et al., 2014 and Lötter, 2014) and obtaining adequate data/information from respondents was a challenge. The researcher had to spend more time explaining the study; finally the respondents gave the required data. More information also had to be obtained in journal articles and studies in related field.

During data collection exercise, some of the respondents declined entirely to taking part in the study explaining that their schedule was too tight and did not have time to attend to the questionnaire or the interview. Some of the respondents, after many reminders, honored their word but a few of them did not respond even after repeated reminders. However, persistent appointments and reminders paid off because the study registered a high response rate.

The findings of the study were limited to agricultural research institutes under KALRO which are funded by the government. International and regional agricultural research institutes would have provided more or different views of the findings and further acted as a benchmark on RDM given their approach on RDM practiced. Unfortunately they did not grant permission to the researcher to undertake research in their institutes.

Lastly, the study was limited by financial and time constrains which restricted the researcher to covering only six (6) agricultural research institutions within KALRO. In this regard, the findings are limited to the 6 mentioned research institutes. However, the researcher felt that the selected agricultural research institutes had provided good case studies given that they covered diverse research disciplines in the agricultural sector.

6.10 Summary of findings mapped to the theoretical models and the research questions

Table 6.1: Summary of findings mapped to the theoretical models and the research questions

Theoretical models	Key variables addressed	Research question	Summary of findings
Community Capability Model (CCM) Framework (Higgins, 2008)	(1)RDM legal framework, (2)RDM policies, (3) RDM regulations,	How does the availability or absence of legal, policy and regulations affect the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes?	(1) Absence of legal framework governing the RDM (2) Inadequate and outdated RDM policies and regulations (3) Absence of RDM unit / department to coordinate functions, activities and service of RDM
Data Curation Centre (DCC) Lifecycle Model (Lyon et al., 2012)	(1) Data capture, (2) Data appraisal, (3) Data description, (4) Data preservation,	How do Kenya's agricultural research institutes capture, appraise, describe, preserve, and make accessible for reuse its research data?	(1)Inadequate modern technologies to capture data in real time (2) Applicability of appraisal and selection policies is poor (3)Absence of standardized format of description (4)Absence of preservation plan and enforcement of the preservation

	(5) Data access, (6) Data reuse (7)Data Sharing		policies to have a harmonized and standardized way of preserving data. (5) Inadequate policies, regulations and mechanisms to facilitate access to agricultural research data (6) Inadequate use and reuse policies to harmonize the way data should be made available for use and reusing. (7) Inadequate policies, strategies and mechanisms of sharing research data
Community Capability Model (CCM) Framework	1)RDM knowledge, (2)RDM skills, (3)RDM Training	What knowledge, skills and training are needed to capture, appraise, describe, preserve, and make accessible for reuse its research data?	(1)Scarcity of RDM knowledge possession by researchers and RDM stakeholders. (2)Scarcity of RDM Skill possessed by researchers and RDM stakeholders. (3)Absences of training policy and strategies

			(4) Limited RDM awareness and advocacy
Community Capability Model (CCM) Framework	ICT(s)	What is the level of ICT preparedness in Kenya's agricultural research institutes for the capture, appraisal, description, preservation, access and reuse of research data?	(1) Inadequate ICT tools and equipment (2)Inadequate use and awareness of ICT infrastructure for RDM. (3) Inadequate RDM security systems
Community Capability Model (CCM) Framework	(1)RDM collaboration, (2)RDM partnership	How do collaborative partnerships influence the capture, appraisal, description, preservation, access and reuse of research data in Kenya's agricultural research institutes?	(1) Inadequate government support in terms of budgetary allocation to do research (2) Inadequate collaborative partnerships on RDM.

Summary of discussion of findings

This Chapter Six discussed and interpreted the findings presented in Chapter Five. The interpretation and discussion of findings covered the main research questions underpinned by the Community Capability Model (CCM) framework (Lyon et al., 2012) and Data Curation Centre (DCC) Lifecycle Model (Higgins, 2008). The chapter therefore, gives meaning and provides implication for the finding presented in chapter Five.

The discussion presented in this chapter has shown that Kenya's agricultural research institutes are partially engaged in RDM based on the stipulations of the ISO 9001:2008 certification standards and the KALRO strategic objectives. The study demonstrated that RDM was partially practiced in the research institutes even if in isolation by the researcher, the heads of research and departments. Despite absence of legal framework in the agricultural research institutes, individual institutes had RDM policies guiding researchers on research data capture, appraisal, description, preservation, access and reuse. Moreover the findings demonstrated that capturing of data was adequately done despite the absence of standardized formats of appraising and describing research data. The lack of coordinated RDM strategies in the research institutes led to loss of data and difficulty in accessing, reusing and sharing research data. Concerning access, reuse and sharing of research data, the finding established that research institutes accessed, reused and shared data modestly and this limited the utilization of RDM. Furthermore, the finding revealed that there was scarcity of RDM knowledge and skill in the agricultural research institutes hence, limiting the utilization of research data in Kenya's agricultural research institutes. The results generally revealed that technical infrastructure for capturing, appraisal, description, preservation, access, sharing and security of data was inadequate to facilitate RDM. It was revealed that collaborative partnership enhanced management, access and sharing of research data. Generally, the findings indicated that support from government, Kenya's agricultural research institutes, and collaborative partners could enhance management, access, reuse and sharing of agricultural output.

It can be concluded that the findings discussed in this chapter corroborates with the finding of Pinfield, Cox and Smith (2014); Cox, Verbaan & Sen (2012); Qin (2013); Mossink, Bijsterbosch, & Nortier, (2013); Grebmer and Spielman (2004); and Lewis (2010) which identified a wider range factors influencing ongoing development in RDM to include polices and legal framework, technology infrastructure, governance, resourcing and skills, and collaborations. These studies adopted qualitative and quantitative epistemologies in examining RDM in research institutions. In contrast, the current study lie in the fact that RDM in Kenya and more so in Africa is given little attention as attested by the limited documentation or publications of the research that is generated by research institutions, leading to limited access of such research, duplication of the research, poor disposal and re-use of the research (Jao et al., 2015; Alila & Atieno, 2006; and Mugata, 2014).

The next chapter (Chapter Seven) provides summary of the findings, conclusion and recommendations.

CHAPTER SEVEN

SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

7.1 Introduction

This chapter presents a summary of the findings, conclusion and recommendations on Research Data Management in Kenya's agricultural research institutes.

The purpose of this study was to examine research data management in Kenya's agricultural research institutes with the view of proposing interventions to improve management, sharing and reuse of agricultural research output. The study was motivated by the fact that in Kenya's agricultural sector RDM is given little attention as attested by the limited documentation or publications of the research that is generated by these institutions, leading to limited access, reuse and sharing of research data, duplication of research, high cost of doing research, potential loss or destruction of data.

The research questions of the study are used as the organizing framework for this chapter under the following subject headings: summary of the findings, conclusion, recommendations and suggested studies for future research.

7.2 Summary of the findings

The summary of findings covers legal, policy and regulations affecting RDM; quality assurance and control measures; capturing, appraising, describing, storage, backup, preserving, accessing, reusing and sharing of research data; knowledge, skills and training among RDM stakeholders; satisfaction with quality of research data services; data literacy programs; level of ICT preparedness for RDM; and collaborative partnerships influencing RDM; data management plan; benefits of RDM; challenges of RDM and recommendations of RDM.

7.2.1 Background information of the respondents

The study revealed that the gender of the respondents were almost the same with male 68(54.8%) and female 56 (45.2%). The study revealed that the majority of the respondents 35(28.2%) were working in research institute A, while 23(18.5%) were working in research institute F. There were 21(16.9%) who were working in research institute C, 16(12.9%) were working in research institute B whereas 15(12.1%) were working in research institute E Only 14(11.3%) were working in research institute D.

The nature of work of the agricultural researchers included among others records management, agricultural pathology, dairy science, agricultural economics, laboratory technology, breeding, zoology/entomology, molecular biotechnology, seed science and technology, among others. The findings further revealed that the majority of respondents were below 50 years old but above 26 years of age, with educational qualifications ranging from Higher diploma holders 11(8.9%), Bachelor's degree 44(35.5%), Master's degree 41(33.1%) and PhD degree holders 21(16.9%) (See section 5.3). Universities generally employed a much higher share of PhD-qualified scientists compared with most National Agricultural Research Institute(s) (NARIs) and other government agencies.

7.2.2 Legal, policy and regulatory framework in RDM

The findings of the study show that that there was no legal framework governing RDM in Kenya's agricultural research institutes (Chapter 5 section 5.4.1) despite the fact that they recognized its imperativeness in the administration of RDM. This meant that researchers, librarians, IT staff and other data professionals were left to rely on experiential knowledge which often led to inconsistency regarding standardization with RDM. Absence of legal framework led to the lack of a clear definition on how research data generated in the agricultural research institutes was managed to ensure continuous appraisal, description, preservation, access, sharing and reuse of the data. Kenya's agricultural research institutes also lacked coordinated responses to RDM. The finding revealed that KALRO Act (No.17 of 2013) does not clearly define how research data generated in the research institutes should be managed to ensure the continued preservation, long-term management, sharing and reuse of agricultural research data. This made RDM in Kenya's agricultural research institute difficult.

The results established that RDM policies and regulations that governed RDM were available in Kenya's agricultural research institutes. RDM policies and regulations covered a wide range of areas related to data curation, sharing, reuse, intellectual property, ethical requirement, ownership of data, security as well as human resource capability, technical infrastructure capability and collaborative partnerships. However, findings revealed that policies on capturing, preservation, access and sharing were partially applied while those covering appraisal, description and reuse policies were rarely applied. It was found that overall, Kenya's agricultural research institutes had a serious problem with regard to enforcing RDM policies and regulations, especially in providing clarity on what was expected in RDM and responsibilities. Moreover, the findings also revealed that generally Kenya's agricultural research institutes did not have standard formats on appraisal, description and preservation of research data leading to poor searching and retrieval, loss of data, and duplication of data.

The overall results showed that there were no policies governing human resource capability for RDM, technical infrastructure for RDM and collaborative partnerships. Absence of policies and regulations in these core areas, results in limited data literacy, advocacy campaigns, inadequate access and sharing of research data and limited accountability in collaborative partnership within as well as outside the research institute.

The study results revealed that Kenya's agricultural research institutes and the researchers owned the right to research data. However, any research data generated from research project belonged (ownership) to the research institute that sponsored the project nevertheless the researcher held the right to data in terms of creation, access and sharing. However, enforcement of the policy was still problematic because most researchers left or disengaged with the research institute without leaving the research data with the institute, resulting in valuable datasets becoming lost or discarded. Besides, the policy did not clearly stipulate the penalties that should be administered to researchers who defied the policy of depositing valuable datasets with the research institute. The RDM policies also failed to indicate where all research data should be deposited and preserved within the research institute, resulting in researchers preserving the data on their personal laptops and individual external drives. This placed agricultural research data at risk of losing data.

Regarding quality assurance and control during data curation, the findings showed that Kenya's agricultural research institutes have complied with quality assurance and control measures during research data capture, appraisal, description, preservation, access and reuse. The attainment of quality assurance and quality control in agricultural research institutes is made possible because KALRO has embraced ISO 9001: 2008. However, the findings established that researchers and other RDM stakeholders inadequately received training on methods for achieving, assessing, or controlling the quality of research data.

7.2.3 Research data management in Kenya's agricultural research institutes

The findings of the study showed that researchers used/adopted diverse methods of capturing research data depending on the research project that was being undertaken. The standard operating procedure(s) (SOPs) generated from ISO 9001:2008 assisted in capturing agricultural research data. However, research data collected in different formats were not managed in a way that could be accessed, reused and shared by other researcher in the research institute or from other research institutes due to poor preservation methods and incompatible formats. It was also found that Kenya's agricultural research data existed in different formats or software versions and was scattered across servers, computers, storage devices or other filing systems, which made it difficult to be accessed, shared and reused.

With regard to appraisal, the findings showed that research institutes' policy and the SOPs were the most preferred tools used to guide the appraisal of research data. However, the use of appraisal and selection and RDM policy was minimal. This was due to inadequate awareness and guidance on the use of appraisal and selection and RDM policy in appraising and selecting research data. Besides appraisal, a key part of RDM is the curation process, which ensures that metadata is available to describe datasets for future use. The overall results showed that different kinds of metadata were used for resource description, discovery, use, presentation, and preservation of research data, however, different researchers adopted different approaches when describing their research datasets depending on the research project they were undertaking. Disparities in the way research datasets were described in various research institutes clearly indicated that there were no harmonized procedures or standards in the way research data was described thus making discoverability, accessibility, reuse and sharing a challenge. In addition, the agricultural research institutes lacked a generalized metadata profile that could be applicable in a standardized model for creating/capturing metadata as well as describing each research datasets.

As with storage of research data, the findings revealed that majority of researchers in Kenya's agricultural research institutes store their research data for the short term in different storage media such as personal laptop, CDs or DVDs, servers or research group data storage. Overall, Kenya's agricultural research institutes do not have standardized storage facilities for storing data in the short term making management, access, reuse, and sharing of research data problematic.

The findings revealed data manager, IT staff, librarian or research unit were responsible for day-to-day management, storage and backup of research data. Overall, Kenya's agricultural research institutes preferred researchers, the data manager and IT department to be responsible for the management, storage and backup of research data. Their different preference was attributed to a lack of awareness regarding who was responsible for day-to-day management, storage and backup of research data, concurrently the RDM policies and regulations did not give guidance on the same.

As to the frequency of making backups, there was no consistency in terms of a clear stipulated period to backup data. The agricultural research institutes also lacked backup strategies for research data. Besides, researchers were not aware of the available technical infrastructure in the agricultural research institutes such as servers in different locations, repositories and secure cloud services to facilitate storage and back up of research data

For preservation of data, findings showed that preservation was taking place majorly on an individual level and minimally at an institutional level. In terms of the length of time for keeping research data, the findings disclosed that researchers had varying times for keeping their research data however, majority 29(23.4%) did not know the length of time for preserving their research data. Moreover, the researchers, researchers in consultation with their researcher supervisor, and funders' requirements were the determinants of research data worth preserving. Furthermore, researchers preferred archiving their research data for long-term preservation as supplementary files attached to the journal articles on publisher's website, others in research institute's data archive server, while some preferred Databank data repositories. On transfer of research datasets for long-term archiving, the findings revealed that researchers preferred transferring research data on physical hard drives and by web-based selected data repositories. Ironically, the library and IT departments played insignificant role in RDM hence missing the expertise of the library and IT department.

On research data access the findings revealed that the absence of legal framework meant that public access to research data varied from one institution to another. As to who could accesses research data, the findings revealed that the majority of people who were currently accessing research data were group researchers, followed by other researchers in the research institute and researchers from other research institutes. There was minimal access to research data by the public.

With regard to data sharing, the findings revealed that researchers shared their research data with members of their research group and trusted external collaborators, their research supervisors, and research sponsors before publication of any papers arising from research data interpretation. Rarely did researchers share research data with the public or everyone, by publishing the data online as indicated by the finding. The researchers did not seem to understand the full potential of sharing data. On methods used in sharing research data, different methods were used with the most preferred as follows: use of external storage device, hard copy/print, through e-mail, depositing them in an institutional repository, and submitting them to a journal to support publication. Inadequate sharing technologies, human resource capability, limited awareness and data literacy limited knowledge transfer beyond scientific community.

As far as reusing data is concerned, the findings revealed that researchers used and reused research data to promote potential new data uses and for encouraging scientific inquiry; re-analysis; promoting innovations, and for reducing duplication to minimize cost of doing research. The findings revealed that lack of advocacy, awareness and data literacy programs, among others, limited the benefits of using and reusing research data.

7.2.4 RDM knowledge, skills and training requirements

The finding of the study showed that Kenya's agricultural research institutes did not have adequate staff knowledgeable in RDM. Of those trained were agricultural researchers but not staff designated as researcher data managers. Consequently, RDM was managed by researchers who did not have much knowledge on technical aspect of RDM. Similarly, research institutes did not have qualified librarians and IT staff knowledgeable in RDM.

Regarding skills in RDM, The findings established that research institutes did not have adequate skilled RDM staff. A few personnel who had RDM skills were researchers, librarians and data managers. As a result, individual researchers managed their own research data despite their limited skills in RDM limiting reuse, and sharing of research data. It was established that researchers and other data professionals had acquired some RDM skill 'on the job', workshops, conferences, seminars, webinars and in house training. Most skills gaps were in the use of institutional repository and developing metadata schema. Therefore training needs were extensive covering these gaps and others such as security and storage of data, data management plans, collaborative partnerships, legal and ethical consideration.

In relation to satisfaction with quality of research data services offered by library and IT department, the findings established that agricultural research institutes have libraries that are rarely used by researchers because: they offered few services related to RDM, information resources available are outdated, absence of knowledgeable and skilled librarians in RDM, poorly equipped with ICT infrastructure, absence of policy mandating the library to support RDM.

Equally, IT departments had very diminutive role to play in RDM but assisted in trouble shooting, repairing researcher's faulty computers, and maintenance of the web page and internet, among others.

Concerning data literacy programs, the findings revealed that researchers and other RDM stakeholders had not undertaken any research data literacy program (or course) to enhance their knowledge and skill in RDM.

7.2.5 Level of ICT preparedness

The findings revealed that agricultural research institutes did not have adequate and up to date ICT tools and equipment to capture, appraise, describe, preserve, access, reuse and share Research Data. Furthermore, the findings revealed that top management of agricultural research institutes did not accord full support to the acquisition of modern ICT tool and equipment used in RDM due to scarce financial resources and recruitment of human resource to operate the ICT tools and equipment. Moreover, the findings also showed that Kenya's agricultural research institutes did not have RDM software to facilitate management, reuse and sharing of agricultural research output.

With regards to data security**,** the findings showed that the agricultural research institutes guarded their research data from plagiarism, copyright and other privacy concerns using physical security, network security, multiple hard drives, cloud for back up and encryption. However, absence of security policy and inadequate awareness on the use of other security methods like security software to protect research data exposed data to a big loss. Meanwhile incompatibility and interoperability of research data security with data repositories, metadata, security systems, data management systems, and search mechanisms made data prone to breaches in the confidentiality of research data.

7.2.6 Collaborative partnerships in RDM

The findings of the study showed that agricultural research institutes were satisfied with collaboration through departmental research groups, collaboration across research group between organizations, and the international collaborations. However, they were dissatisfied with collaborations organized at national level due to inadequate government support. Similarly, absences of legal framework, outdated RDM policies and unwillingness of research institute management to initiate a working model for collaboration at national level hampered a vibrant and effective RDM at national level.

The findings further established that the research institutes were satisfied with individual researchers occasionally collaborating outside their discipline, interactions across discipline collaborating through joint conferences or publications, and formal collaboration between research groups from different disciplines. Nonetheless, absences of comprehensive and articulate collaborative partnership policies and strategies, and legal frameworks limited the nurturing of collaborative culture on accessing and sharing knowledge. Concerning collaboration with the public, the findings established that inadequate access and sharing policies limited collaborations with the public, which in turn impeded the use and reuse of research data. However, there was satisfaction in participative media programs and other forums organized to engage the public such as trade fairs and agricultural shows, among others.

The findings showed that agricultural research institutes were satisfied with public-private partnerships. Through these partnerships, the research institute were able to access and share research data, share research tools and equipment, repositories, human expertise, share statistical data, learn new skills, build and empower those without resources and ICT infrastructure. They also benefitted through joint publications, networking, joint innovation, and patenting.

7.2.7 Benefits of RDM

The study found that research institutes were inadequately benefiting from RDM due to the absence of legal framework, outdated RDM policies and regulation, absences of RDM coordinating unit, lack of trained RDM staff among others. Benefits of RDM that would include facilitation of access, reuse and sharing of research, increasing research efficiency, enhancing research visibility and more were not optimized.

7.2.8 Challenges encountered in RDM and amelioration strategies

The findings established that agricultural research institutes were struggling with RDM challenges such as: an absence of RDM legal framework and outdated policies and regulation; absence of qualified staff with RDM knowledge and skill; inadequate data literacy and advocacy; inadequate technical infrastructure; inadequate funding to facilitate researchers to do research; absence of collaborative research data access and sharing; minimal collaborations among researchers, absence of RDM units, among others. To address these challenges, the following interventions were proposed: enactment of RDM legal framework and increased funding for research; improvement of ICTs infrastructure for RDM, skills development in RDM; promulgation of RDM policies and regulation; and large scale collaborative partnerships.

7.3 Conclusion

This section provided conclusions based on the major findings of the study. The conclusions were drawn in the order in which the research questions were stated in chapter one**.**

The findings pointed out the need for RDM legal framework to reinforce the establishment of RDM as a statutory responsibility. Legal framework would set up structures and give directions on the activities, functions and services of RDM. Such RDM legal frameworks would need to be comprehensive, articulate, coherent and consistent covering all stages of data curation lifecycle, sharing, reuse, including human resource capability, technical infrastructure and collaborative partnerships.

The findings seem to point to the need for RDM policies and regulations that are in line with the core mandate and statutory responsibilities of agriculture research institutes. Agricultural research institutes demonstrated capability for practicing RDM with regard to the creation, dissemination, embodiment of products, services and systems.

The findings suggest that research data capture, appraisal, description, preservation, accessibility, reuse and sharing in the agricultural research institutes were the core functions, activities and drivers of RDM. There was clearly growing impetus in data curation lifecycle as revealed by growing interest in long-term preservation, access, reuse and sharing of research data.

The findings suggested that research data in the research institutes were not professionally managed from creation to sharing as envisaged by DCC lifecycle model and CCM framework due to a lack of human resources with RDM knowledge and skills who have competencies in data literacy; intellectual property; scholarly communication; cyberinfrastructure, metadata, among others.

The level of ICT(s) infrastructure preparedness with regard to RDM was rated between 50% in some institutes and 80% as the highest level of implementation in the agricultural research institutes. This implies that technical infrastructure is partially in place but the application of it for

RDM is minimal. The following technical infrastructure was in place among others; institutional repository, emails, Internet, web portals, and more. Limited government support and absence of RDM legal framework has affected acquisition of ICTs tools and equipment used in RDM. The findings also seem to suggest that collaborative partnerships were generally inadequate and unsatisfactory.

Overall the findings suggest that although Kenya government attaches great importance to the agricultural sector, as demonstrated by the establishment of KALRO vide the KALRO Act (N0.17 of 2013) to coordinate agricultural research in the country, the Act does not clearly define how research data generated in the research institutes should be managed to ensure the continued preservation, long-term access, sharing and reuse of the data. Consequently, there were several weaknesses in RDM in agricultural research institutes. These weaknesses include: absence of RDM legal framework; inadequate and outdated RDM policies and regulations; absence of RDM unit/department to coordinate functions, activities and service of RDM; inadequate data literacy, limited awareness and RDM advocacy; inadequate RDM security systems; absence of RDM guidelines on standardization; inadequate technical infrastructure; inadequate knowledge, skills and training on RDM; and inadequate collaborative partnerships on RDM.

7.4 Recommendations

The study has discussed various issues on RDM in Kenya's agricultural research institutes and established that RDM in agricultural research institutes faced scores of challenges. Based on the finding of the study, the interpretation and conclusion adduced above, the recommendations are proffered in section 7.4.1 to 7.4.6.

7.4.1 Legal, policy and regulatory framework

The study revealed that agricultural research institutes did not have RDM legal framework. In the same vein, RDM policies and regulation were inadequate and outdated.

Recommendation 1: ***Data Governance***: - A formal data governance structure is recommended in order to address the wide variety of data issues in agricultural research institutes. A robust data governance framework will provide the structure and institutional oversight necessary to establish a culture of data fluency across the institutes (Flory and Walker, 2015). Such governance program includes a governing council, set of procedures, and a plan to execute that procedure. Flory and Walker (2015); Brown, Bruce and Kernohan (2015); and RECODE Project Consortium (2014) assert that under data governance RDM legal and policy frameworks must be developed, processes must be defined concerning how the data is to be captured appraised, described, preserved, backed up, accessed, shared, reused and secured. In addition, standards and procedures must be developed that define how data is to be used, by whom, and for what purposes. The data governance should also put into place a set of controls and audit procedures to ensure ongoing compliance with RDM legal and policy framework.

Recommendation 2: *Legal framework*:- The study strongly recommends that the government should enact legislation to give research institutes statutory mandate to capture, appraise, describe, preserve, access, reuse and share research data in all research institutes as per DCC lifecycle model (Higgins, 2008) and CCM framework (Lyon et al., 2012). Such an Act should clearly define how research data generated in the research institutes should be managed to ensure the continued preservation, long-term access, sharing and reuse of research data, research data ownership, and intellectual property rights among others. The legal framework will give effect and provide a roadmap on issues like human resource capability, technical infrastructure, collaborative partnerships among other issues related to RDM.

Recommendation 3: *RDM policy and regulations*: - It is recommended that Kenya's agricultural research institutes should consider revising RDM policies and regulations to include current trends in RDM (e-research, cyberinfrastructure); mechanisms of enforcing policies; administering penalties; functions, activities, service and responsibilities among other. The RDM policies and regulations should be aligned with legal framework in order to create meaningful and vibrant RDM in Kenya's agricultural research institutes. In addition, there should be mechanisms in place that enable regular review of the policies and a team to foresee its implementation. RDM policies and regulations should be clear, comprehensive, consistent, coherent and articulate mapping out on every RDM functions, drivers, relationships and influences of RDM.

7.4.2 Data capture, appraisal, description, preservation, accessibility and reuse

The study revealed many challenges were faced in the capture, appraised, description preservation, access; reuse and sharing in Kenya's agricultural research institutes.

Recommendation 4: *RDM Unit/department*: - Regarding RDM unit, the study recommends the establishment of RDM unit/department to oversee the functions, activities, roles, services, responsibilities and coordination of research data capture, appraisal, description, preservation, accessibility, reuse and sharing in all agricultural research institutes. The same unit will be responsible for laying down strategies and mechanisms for RDM; implementation and reinforcement of RDM policies and regulation; running advocacy campaigns and creating, awareness about RDM. King Abdullah Medical City (2017) add that RDM unit/department should be mandated with developing collection instruments, designing database, cleaning and validating data, and helping researchers to collaborate better nationally and internationally. The unit will also be responsible for portal services, metadata standards and management. Johnsson and Ahlfeldt (2015); Lewis (2010); Flores et al., (2015) are also of the opinion that RDM unit/department should responsible for data curation; data discovery, access, reuse and sharing of research data. Similarly, the unit/department should enact ethical procedures or issues to foresee management of research data.

7.4.3 RDM knowledge, skills and training requirements

The study findings revealed that agricultural research institutes did not have human resource capability specifically for RDM. Researchers and heads of research acted as the custodians of RDM.

Recommendation 5: *Human resource capability for RDM*: - It is recommended that agricultural research institutes should recruit personnel possessing knowledge and skills for RDM. RDM staff would have the responsibility of research data capture, appraisal description, preservation, access, reuse and sharing. Johnsson and Ahlfeldt (2015) emphasized the need to recruit staff that possesses RDM competencies such as resource documentation, storage and description; knowledge organization and subject/discipline structures; and metadata management. In addition, Brown, Bruce and Kernohan (2015) assert that RDM staff should possess skills in policy development, business analysis, advocacy, project management, metadata, data archiving and preservation, among others.

Recommendation 6: *Advocacy, awareness and data literacy programs*: - The study recommends the establishment of advocacy and data literacy programs to enable researchers and other RDM stakeholders to understand the functions, activities and benefits of RDM. Data literacy and awareness programs play a crucial role in RDM because they develop in researchers and other RDM shareholders an understanding of the way in which research data is generated, how they need to describe it to facilitate future retrieval, how to preserve, access, share and reuse research data (Lewis, 2010 and Flores et al., 215). Advocacy with regard to resource allocation should also be addressed by the government and management of agricultural institute.

Recommendation 7: ***Library and IT department role in RDM***: - The study recommends that library and IT departments should be involved in RDM to address challenges and offer expertise on data curation, reuse and sharing of research data. Recruiting RDM librarians for the purpose of designing data management support would be helpful to foster an interoperable infrastructure for data access, discovery and sharing, development of data management plans, institutional repository services, intellectual property rights advice and development of a data management portal. Flores et al., (2015); Corrall et al., (2013); and Cox and Pinfield, (2014) point out that when considering the library and IT department role in RDM, common themes include: activities associated with conducting RDM needs assessment in user community; advocacy, awareness, and training; advisory services; data repository development; backup; storage; metadata among others.

Recommendation 8***: Incentives:*** **-** The study recommends incentives for researchers, as they need explicit, meaningful rewards for engaging effectively with RDM, either through the Research Excellence Framework (REF) or in the form of career progression within their agricultural research institutes. Brown, Bruce and Kernohan (2015) assert that agricultural research institutes and their RDM staff should find practical ways to shift from compliance to professional rewards for researchers.

7.4.4 Level of ICT preparedness

The study showed that the level of ICT preparedness in Kenya's agricultural research institutes were not adequate.

Recommendation 9: ***Technical infrastructure***: - The study therefore recommends that the agricultural research institutes should fully adopt ICTs for RDM to enhance the security, accessibility, efficiency, reliability and responsiveness of RDM. ICTs should facilitate all functions, roles, activities and services of research data capture, appraisal, description, preservation, access, reuse and sharing. The agricultural research institutes should invest in modern technologies such as data repositories, high-speed gateways, fiber optic, web 2.0, multimedia technologies, among others. Mechanisms should be developed to enhance compatibility of research data security with data repositories, metadata, security systems, data management systems, and search mechanism.

7.4.5 Collaborative partnerships

The study findings revealed that although collaborative partnerships were seen as an important component in RDM in the agricultural research institutes, Kenya government had not fully supported collaborations at national level.

Recommendation 10: ***Government support***: - The study recommends that the government, through legislative and policy framework, in enhancing collaborative partnerships, should support RDM. Through budgetary allocation, researchers are able to do research and generate agricultural output to be managed, reused and shared with other collaborators. Kahn et al., (2014) posit that government support in terms of planning and resource allocation plays a fundamental role in carrying out research, which in turn generates research data for curation, access, reuse, and sharing among collaborating research institutes, groups or individuals.

Recommendation 11: ***Data management plan***: - The study recommends that the existing SOPs should be standardized in all agricultural research institutes and has it changed to data management plan of which it will elaborate and focus more on RDM. Fitzgerald, Pappalardo and Austin (2008) opined that DMP addresses how data is collected, appraised, described, preserved, accessed, shared and disseminated. It also addresses data ownership, data security and quality assurance.

7.5. Contribution and originality of the study

For research to be of value, it should address issues that are important to a particular society, community or institution (Wassenaar, 2006). The research questions addressed in this study were of value to government and non-governmental organizations, international, regional and national agricultural research institutes, other research institutes outside agriculture, policy-makers, researchers, extension workers, and farmers in Kenya in relation to RDM.

From a legal and policy perspective, the findings have the potential to influence the formulation of RDM legal and policy framework for Kenya's agricultural research institutes. The findings provide legal and policy directions to government policy-makers, management of Kenya's agricultural institutes, researchers, RDM staff, and extension services in developing RDM legal and policy frameworks concerning research data capture, appraisal, description, preservation, access, share, and reuse; human resource capability; technical infrastructure; collaborative partnerships, security, quality assurance, among others. From a pragmatic perspective, the findings uncovered and created awareness about the importance of RDM in agricultural research institutes and in this regard the recommendation adduced by the study can be espoused to improve or implement RDM in agricultural research institutes in order to promote management, access, sharing and reuse of agricultural research output. In the wake of embracing RDM recommendations, pragmatic benefits will be to: avoid duplication of research projects, reduce costs of doing research, meet research funders requirements, retrieval comparison and co-analysis of data from multiple sources that could lead to powerful insights among others.

For theory, the study contributes to the domain of knowledge and literature, especially in the context of RDM in Kenya. Moreover, extant literature on RDM has tended to concentrated on libraries, and related ICTs with little regard if any to agricultural research sector (Lewis, 2010; Pinfield, Cox and Smith, 2014). The current study therefore examines RDM in Kenya's agricultural research institutes thus contributing new knowledge and literature, new and innovative views of discussing theories underpinning the study and opening up new areas for research. The study recommended the enactment of RDM legal framework, revision and re-alignment of RDM policies and regulations within the legal framework, taking into account its enforcement, establishment of RDM unit with the mandate of RDM, among others (see chapter 7 section 7.4).

Furthermore, literature reviewed revealed that RDM in Kenya's agricultural sector is given little attention as attested by the limited documentation or publications of the research that is generated by these institutions, leading to poor management, limited access of such research data, duplication of the research, poor sharing and reuse of the research data (Jao et al., 2015; Family Health International-Kenya, 2005; The World Agroforestry Centre, 2012; Alila and Atieno, 2006; Mugata, 2014). Studies such as the current one contribute frameworks particular of specific to developing countries by providing empirical evidence of specific challenges that such countries contend with. Moreover, few empirical studies on RDM have been done in Africa and particularly Kenya. The current study is therefore significant in contributing to the scholarly research and literature on RDM in developing countries such as Kenya.

Table 6.1 presents a summary of findings mapped to the theoretical models underpinning the study, attributes of the models and the research questions.

7.6 Suggestion for further research

The present study investigated research data management in Kenya's agricultural research institutes with the view to proposing interventions to improve management, sharing and reuse of agricultural research output. The study investigated six (6) of Kenya's agricultural research institutes (see section 4.6). However, there are other research institutes in Kenya, which focus on research in other fields that generate massive research data. These include KEFRI, Kenya Medical Research Institute (KEMRI), and Kenya Industrial Research Institute (KIRDI) among others. The current study recommends research to be extended to cover these research institutes.

Furthermore, the current study was conducted in government-sponsored organizations in Kenya. Future research could be extended to international research institutes in Kenya in the country for bench marking purposes. The international research institutes in Kenya include ILRI, ICRAF, The International Centre of Insect Physiology and Ecology (ICIPE), among others.

REFERENCE

Agricultural Information Management Standards (AIMS). (2013). G8 Science Ministers Endorse the need to Increase Access to Publicly-Funded Research. [Online]. Available at: http://aims.fao.org/activity/blog/g8-science-ministers-endorse-need-increase-access-publicly-funded-research [Accessed 3 September 2016].

Agricultural Information Management Standards (AIMS). (2017) Research Data. [Online]. Available at: http://aims.fao.org/fr/research-data [Accessed 29 March 2017].

Alila, P. and Atieno, R. (2006). Institute of Development Studies Agricultural Policy in Kenya: Issues and Processes. *A Paper for the Future Agricultures Consortium Workshop, Institute of Development Studies, 20-22 March 2006*. [Online]. Available at: http://www.fao.org/fileadmin/user_upload/fsn/docs/Ag_policy_Kenya.pdf [Accessed 25 April 2016].

American Psychological Association. (2015). Data Sharing: Principles and Considerations for Policy Development. [Online]. Available at: https://www.apa.org/science/leadership/bsa/data-sharing-report.pdf [Accessed 5 October 2016].

Amorim, R.C., Castro, J.A., Silva, J. R. and Ribeiro, C. (2015). A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential. [Online]. Available at: http://link.springer.com/chapter/10.1007/978-3-319-16486-1_10#page-1 [Accessed 13 May 2016].

Anderson, W.L. (2004). Some Challenges and Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data. *Data Science Journal,* 3, 191–202. [Online]. Available at: http://www.jstage.jst.go.jp/article/dsj/3/0/191/_pdf [Accessed 31 March 2016].

Antell, K., Foote, J.B., Turner, J. and Shults, B. (2014). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. [Online]. Available at: http://crl.acrl.org/content/75/4/557.full.pdf+html [Accessed 20 April 2016].

Antelman, K. (2004). Do Open-Access Articles have a Greater Research Impact? *College Research Libraries,* 65(5), 372-382.

Arzberge, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P. and Wouters, P. (2014). Promoting Access to Public Research Data for Scientific, Economic and Social Development. *Data Science Journal,* 3(29), 135-152.

Ashurst, M. and Mbithi, S. (2010). Why Africa can make it Big in Agriculture. [Online]. Available at: http://africaresearchinstitute.org/newsite/wp-content/uploads/2013/03/Why-Africa-can-make-it-big-in-agriculture.pdf [Accessed 3 October 2016].

Association of Research Libraries. (2014). Scholarly Communication. [Online]. Available at: http://www.arl.org/focus-areas/scholarly-communication [Accessed 20 October 2016].

Australian Catholic University. (2017). Research Data Management Policy. [Online]. Available at: https://www.acu.edu.au/policy/research/general_policies/research_data_management_policic [Accessed 24 August 2017].

Australian Government. (n.d). The Australian Research Data Infrastructure Strategy. [Online]. Available at: https://docs.education.gov.au/system/files/doc/other/the_australian_research_data_infrastructure_strategy.pdf [Accessed 5 August 2016].

Australian National Data Service. (n.d(a)). Data Re-use. [Online]. Available at: http://www.ands.org.au/working-with-data/enabling-data-re-use/data-re-use [Accessed 27 October 2016].

Australian National Data Service. (n.d(b)). Metadata. [Online]. Available at: http://www.ands.org.au/working-with-data/metadata [Accessed 27 October 2016].

Austrian National Data Services (ANDS) Technical Working Group. (2007). Towards the Australian Data Commons: a Proposal for an Australia National Data Service. Department of Education, Science and Training. [Online]. Available at: http://www.ands.org.au/__data/assets/pdf_file/0009/386982/ardc_eif_annual_report_2011-12.pdf [Accessed 26 May 2016].

Awuor, F. M., Rabah, K. and Kimeli, K. (2013). E-Agriculture Framework: Modeling Stakeholders' Competing and Conflicting Interests. *International journal of Agricultural Policy and Research*, 1(6), 172-179.

Babbie, E. R. (2001). *The Practice of Social Research*. (9th ed.). Belmont, CA (USA): Wadsworth/Thomson Learning.

Babbie, E. R. & Mouton, J. (2001). The Practice of Social Research. (8th ed.). England, UK: Oxford University Press.

Ball, A. (2012). Review of Data Management Lifecycle Models. [Online]. Available at: http://opus.bath.ac.uk/28587 [Accessed 20 July 2016].

Barber, D., and Zauha, J. (1995). Scientific Data and Social Science Data Libraries. *IASSIST Quarterly*, 19(4), 5-6.

Becker, H. (1998). *Tricks of the Trade: How to think about your Research while you're Doing It.* Chicago: University of Chicago Press.

Beintema, N. (2015). African States should Double Agriculture Research Spending. Standard Newspaper (Standard digital. [Online]. Available at: http://www.standardmedia.co.ke/mobile/article/2000142775 [Accessed 25 April 2016].

Beintema, N. and Stads, G. (2017). A Comprehensive Overview of Investments and Human Resource Capacity in African Agricultural Research. [Online]. Available at: https://www.asti.cgiar.org/sites/default/files/pdf/SSA-Synthesis-Report-2017.pdf [Accessed 28 June 2017].

Berlin Declaration. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. [Online]. Available at: http://www.fu-berlin.de/sites/open_access/Veranstaltungen/oa_berlin/poster/Berlin-Dclaration_Simone-Rieger_MPIWG.pdf [Accessed 25 April 2016].

Bernath, U. and Vidal, M. (2007). The Theories and the Theorists: Why Theory is Important for Research. [Online]. Available at: http://www.cairn.info/revue-distances-et-savoirs-2007-3-page-427.html [Accessed 30 June 2016].

Bhattacherjee, A. (2012). Social Science Research: Principles, Methods, and Practices. [Online]. Available at: http://scholarcommons.usf.edu/oa_textbooks/3 [Accessed 30 July 2016].

Biddick, M. (2012). Research: The Big Data Management Challenges. Information Week. [Online]. Available at: http://reports.informationweek.com/abstract/81/8766/business-intelligence-andinformation-management/research-the-big-data-management-challenge.html [Accessed 19 September 2017].

Bigagli, L., Sveinsdottir, T., Wessel, B., Smallwood, R., Linde, P., and Sondervan, J. (2013). Policy Recommendations for Open Access to Research Data in Europe. [Online]. Available at: http://policy.recodeproject.eu/assets/recode-handbook.pdf [Accessed 3 April 2016].

Biotechnology Research Institute (BRI). (2015). Human Resource Database- Lodged at BRI. Unpublished.

Bloomberg, L.D. and Volpe, M. (2008). *Completing your Qualitative Dissertation: A Road Map from Beginning to End.* London: Sage.

Bogdan, R.C. and Biklen, S.K. (2006). *Qualitative Research for Education: An Introductory Theory and Methods.* 5th (eds.). Needham Heights, MA: Allyn and Bacon.

Bohémier, K.T., Atwood, A., Kuehn, A. and Qin, J. (2011). A Content Analysis of Institutional Data Policies. Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL'11), June 13-17, 2011, Ottawa, Canada, pp. 409-410.

Borgman, C.L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*. 63, 1059-1078.

Boston University. (n.d). Importance of Data Management. [Online]. Available at: https://www.bu.edu/datamanagement/background/importance/ [Accessed 29 June 2017].

Boston University Libraries. (n.d). What is Research Data? [Online]. Available at: http://www.bu.edu/datamanagement/background/what ids data/ [Accessed 9 March 2017].

Bowen, G.A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal,* 9(2), 27-40.

Bracke, M.S. (2011). Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data. *Journal of Agricultural and Food Information,* 12 (1), 65-74.

British Standards Institute. (2009). *Secure Destruction of Confidential Material*, Code of Practice BS EN 15713. [Online]. Available at: http://www.isoqsltd.com/iso-standards/bs-en-15713-secure-destruction-confidential-material/ [Accessed 3 July 2016].

Brown, S., Bruce, R. and Kernohan, D. (2015). Directions for Research Data Management in UK Universities. [Online]. Available at: https://www.fosteropenscience.eu/sites/default/files/pdf/1240.pdf [Accessed 19 September 2017].

Bryman, A. (2006). Integrating Quantitative and Qualitative Research: How is it done? *Qualitative Research*, 6(1), 97-113.

Bryman, A. (2012). *Social Research Methods*. 5th (ed.). London: Oxford University Press.

Bryman, A., Teevan, J. J. and Bell, E. (2009). *Social Research Methods*. 2nd Canadian (ed.). Canada: Oxford University Press.

Burton, L.D. (2010). *Agriculture: Fundamentals and Applications*. 5th (ed.). New York: Delmar.

Bush, V. (2009). Research Data Management. [Online]. Available at: https://net.educause.edu/ir/library/pdf/ers0908/rs/ers09087.pdf. [Accessed 14 May 2016].

Buys, C.M. and Shaw, P.L. (2015). Data Management Practices Across an Institution: Survey and Report. *Journal of Librarianship and Scholarly Communication*, 3(2), 1-15.

Campbell, E., Clarridg, B.R., Gokhale, M., Birenbaum, L. and Hilgartner, S. et al. (2002). Data Withholding in Academic Genetic: Evidence from a National Survey. *Journal of the American Medical Association,* 287(4), 473-480.

Carlson, J. and Leiter, D. (2009). Addressing Researcher's needs through the Data Curation Profile. *In Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL '09), 365-366.

Carrick, A. (2014). Research Data Management: Guidelines and Planning for Researchers. [Online]. Available at: https://www.swinburne.edu.au/media/swinburneeduau/research/docs/pdfs/Research_Data_Management_Guidelines.pdf [Accessed 22 May 2017].

Check, J. and Schutt, R. T. (2012). *Research Methods in Education*. Los Angeles: sage.

Cherryholmes, C.H. (1992). Notes on Pragmatism and Scientific Realism. *Educational Researcher*, 21(6), 13-17.

Chiliswa, Z. and Mutuku, L. (2015). Building Open Data Infrastructure and Strategies for effective Citizen Engagement. [Online]. Available at: http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper13.pdf [Accessed 29 November 2016].

Chinyemba, A. (2005). Managing Records at Higher Education Institutions: A Case Study of the University of KwaZulu-Natal, Pietermaritzburg Campus. *South African Journal of Information Management,* 7 (1), 1-19.

Chisenga, J. (2012). *Opening Access to Agricultural Sciences and Technical Information in Ghana, Kenya and Zambia.* Presented at SCECSAL XXth Conference hosted by Kenya Library Association (KLA) on 4th-8th July, 2012. Venue LAICO REGENCY HOTEL, Nairobi, Kenya. [Online]. Available at: https://core.ac.uk/download/pdf/11890652.pdf [Accessed 4 March 2017].

Chisenga, J., Kedemi, R., Sam, J., Rudgard, S. and Martin, F. (2011). Disseminating the Output of Agricultural Research in Africa. An International Journal for Information Specialists in Agriculture, Natural Resources, and the Environment, 4(1), 40-47.

Coffee Research Institute (CRI). (2015). Human Resource Database- Lodged at CRI. Unpublished.

Cohen, L., Manion, L. and Marrison, K. (2007). *Research Methods in Education.* 6th (ed.). London: Routledge.

Coherence in Information for Agricultural Research for Development (CIARD). (2012). Case Studies: Kenya Agricultural Information Network (KAINET). [Online]. Available at: http://www.ciard.info/sites/default/files/RAF%20(Kenya).pdf [Accessed 30 March 2016].

Cooper, H. M. (2011). *Research Synthesis and Meta-Analysis: A Step by Step Approach*. 4th (ed). London: Sage.

Corbin, J. and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* 3th (ed.). Thousand Oaks, CA: Sage.

Cornell University Library. (n.d). Metadata and Describing Data. [Online]. Available at: http://data.research.cornell.edu/content/writing-metadata [Accessed 6 June 2016].

Corrall, S., Kennan, M. A. and Afzal, W. (2013). Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, 61(3), 636-674.

Council on Library and Information Resources. (2013). Research Data Management: Principles, Practices and Prospects. [Online]. Available at: https://www.clir.org/pubs/reports/pub160/pub160.pdf [Accessed 3 July 2017].

Cox, A.M. and Pinfield, S. (2014). Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316.

Cox, A.M., Verbaan, E. and Sen, B. (2012a). Rising to the Research Data Management Challenges. [Online]. Available at: http://www.sconul.ac.uk/sites/default/files/documents/13_13.pdf [Accessed 1 July 2017].

Cox, A., Verbaan, E. and Sen, B. (2012b). Upskilling Liason Librarians for Research Data Management. [Online]. Available at: http://www.ariadne.ac.uk/issue70/cox-et-al [Accessed 27 October 2016].

Creamer, A., Morales, M.E. and Crespo, J. (2012). An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England. *Journal of eScience Librarianship*, 1(1), 18-26.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative and mixed Methods Approaches*. 2nd (ed.). Thousand Oaks, London: Sage Publications.

Creswell, J. W. (2008). *Mixed Methods Research: Design and Procedures*. Los Angeles: Sage.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches.* 3rd (ed.). Los Angeles: Sage.

Creswell, J. W. (2012). *Education Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. 4th (ed.). Boston, MA: Pearson.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 3th (ed.). Thousand Oaks, California: Sage.

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and mixed Methods Approaches*. 4th ed. Thousand Oaks, California: Sage.

Creswell, J. W. and Piano Clark, V. (2007). *Designing and Conducting Mixed Methods Research.* Thousand Oaks, California: Sage.

Creswell, J. W. and Plano Clark, V.L. (2011). *Designing and Conducting Mixed Methods Research*. 2nd ed. Los Angeles: Sage.

Crownston, K. and Qin, J. (2012). A Capability Maturity Model for Scientific Data Management: Evidence from the Literature. [Online]. Available at: https://www.asis.org/asist2011/proceedings/submissions/36_FINAL_SUBMISSION.pdf [Accessed 31 May 2016].

Dairy Research Institute (DRI). (2015). Human resource database-Lodged at DRI. Unpublished.

Data Curation Centre. (2004a). History of the DCC. [Online]. Available at: http://www.dcc.ac.uk/about-us/history-dcc/history-dcc [Accessed 1 July 2016].

Data Curation Centre. (2004b). Planning Preservation. [Online]. Available at: http://www.dcc.ac.uk/digital-curation/planning-preservation [Accessed 4 July 2016].

Davies, D. and Dodd, J. (2002). Qualitative Research and the Question of Rigor. *Qualitative Health Research*, 12 (2), 279-289.

Decman, M. and Vintar, M. (2013). A Possible Solution for Digital Preservation of e-Government: A Centralised Repository within a Cloud Computing Framework. *Aslip Proceedings: New Information Perspectives,* 65, 406-424.

De Los Reyes, A. and Kazdin, A. E. (2008). When the Evidence Says 'yes, no and may be so': Attending to and Interpreting Inconsistence Findings Among Evidence-based Interventions. *Current Directions in Psychological Science,* 17(1), 47-51.

Denscombe, M. (2007). The Good Research Guide for Small-Scale Social Research Project. 3rd (ed.). Maidenhead: Open University Press.

Denton Declaration. (2012). Open Access @ UNT. University of North Texas. [Online]. Available at: https://openaccess.unt.edu/denton-declaration [Accessed 2 December 2016].

Denzin, N.K. and Lincoln, Y.S. (2005). Introduction: The Discipline and Practice of Qualitative Research. In N.K. Denzin and Y.S. Lincoln (Eds.). *The Sage Handbook of Qualitative Research*. 3rd (ed.). 1-32. Thousand Oaks, CA: sage.

Devex (2017). Kenya Agricultural and Livestock Research Organization (KALRO). [Online]. Available at: https://www.devex.com/organizations/kenya-agricultural-and-livestock-research-organization-kalro-22923 [Accessed 29 March 2017].

Donnelly, M. (2012). Data Management Plans and Planning. In G. Pryor (Ed.). *Managing Research Data.* London: Facet Publishing.

Dora, M. and Kumar, H. A. (2015). Managing Research Data in Academic Institutions: Role of Libraries, 2015. In CALIBER, 2015, Shimla, March 12-+14, 2015 [Conference paper]. [Online]. Available at: http://eprints.rclis.org/24911/1/50.pdf [Accessed 18 September 2017].

Eaker, C. (2016). Selection and Appraisal of Digital Research datasets. In L. Kellam and K. Thompson (Eds.). *Databrarianship: The Academic data Librarian in Theory and Practice. American Library Association.* [Online]. Available at: http://trace.tennessee.edu/cgi/viewcontent.cgi?article=1018andcontext=utk_libpub [Accessed 4 October 2016].

Edge, P., Rudgard, S. Maru, A. and Porcari, E. (2011). Enhancing Communication of Agricultural Research Outputs Using Information and Communication Technology. *An International Journal for Information Specialists in Agriculture, Natural Resources, and the Environment*, 4(1), 32-39.

Edmonds, W.A. and Kennedy, T.D. (2013). *An Applied Reference Guide to Research Designs*. Chichester: Sage.

Eisenhardt, K.M. and Graebner, M.E. (2007). Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal*, 50 (1), 25-32.

Eisner, E.W. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Education Practice.* New York: Macmillan Publishing Company.

EnablingOpenScholarship. (n.d) Open Access Institutional Repositories: A Brief Paper. [Online]. Available at: http://www.openscholarship.org/upload/docs/application/pdf/2009-09/open_access_institutional_repositories.pdf [Accessed 1 July 2016].

Enke, N., Thessen, A., Kerstin, B., Bendix, J. and Seeger, B. et al (2012). The User's View on Biodiversity Data Sharing-Investigating Facts of Acceptance and Requirements to Realize a Sustainable Use of Research Data. *Ecological Informatics*, 11, 25-33.

Ennis, C.D. (1999). The Theoretical Framework: The Central Piece in the Research Plan. *Journal of Teaching in Physical Education*, 18, 129-140.

Erway, R. (2013). Starting the Conversation: University-Wide Research Data Management Policy. Dublin, Ohio: OCLC Research. [Online]. Available at: http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf [Accessed 1 May 2016].

Erway, R. and Rinehart, A. (2016). *If You Build It, Will They Fund? Making Research Data Management Sustainable*. Dublin, Ohio: OCLC Research.

European Commission. (2016). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020: Version 2.1. [Online]. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [Accessed 20 April 2016].

European Strategy Forum on Research Infrastructures. (2009). e-IRG Report on Data Management: Data Management Task Force. [Online]. Available at: https://ec.europa.eu/research/infrastructures/pdf/esfri/publications/esfri_e_irg_report_data_management_december_2009_en.pdf [Accessed 14 May 2016].

Fain, J.A. (2013). *Reading, Understanding and Applying Nursing Research*. Philadelphia: F.A. David Company.

Family Health International-Kenya. (2005). Strengthening Research and Data Management: New efforts at the Kenya Division of Reproductive Health. [Online]. Available at: http://pdf.usaid.gov/pdf_docs/Pnadk371.pdf [Accessed 29 March 2016].

Farnel, S. and Shiri, A. (2014). Metadata for Research Data: Current Practices and Trends. [Online]. Available at: http://dcpapers.dublincore.org/pubs/article/view/3714/1937 [Accessed 20 June 2017].

Fary, M and Owen, K. (2013). Developing an Institutional Research Data Management Plan Service. [Online]. Available at: http https://net.educause.edu/ir/library/pdf/acti1301.pdf [Accessed 10 March 2016].

Fellous-Sigrist, M. (2015). What are Research Data? [Online]. Available at: https://www.ucl.ac.uk/tobedeleted-sites/dis-old-June16/research/RDM_presentation.pdf [Accessed 1 March 2017].

Fernihough, S. (2011). E-Research: An Implementation Framework for South African Organization (Unpublished Master's Research Report, University of South Africa) [Online]. Available at: http://hdl.handle.net/10500/4474 [Accessed 26 May 2016].

Fitzgerald, A. and Pappalardo, K. (2007). Building the Infrastructure for Data Access and Reuse in Collaborative Research. [Online]. Available at: https://eprints.qut.edu.au/8865/1/8865.pdf [Accessed 4 June 2017].

Fitzgerald, A., Pappalardo, K. and Austin, A. (2008). Practical Data Management: A Legal and Policy Guide. [Online]. Available at: https://eprints.qut.edu.au/14923/1/Microsoft_Word_-_Practical_Data_Management_-_A_Legal_and_Policy_Guide_doc.pdf [Accessed 5 June 2017].

Flores, J.R., Brodeur, J.J., Daniels, M.G., Nicholls, N. and Turnator, E. (2015). Libraries and the Research Data Management Landscape. [Online]. Available at: http://www.clir.org/pubs/reports/pub167/RDM.pdf [Accessed 3 December 2015].

Flory, S. and Walker, A. (2015). idata: Data Management for Higher Education. [Online]. Available at: https://www.csuci.edu/irpe/dgc-documents/march-2016/idata-recommendations-for-ci.pdf [Accessed 25 September 2017].

Food and Agricultural Organization (United Nations) (FAO). (2002). Virtual Extension, Research and Communication Network (VERCON), Research, Extension and Training Division (SDR) and World Agricultural Information Centre (WAICENT). [Online]. Available at: ftp://ftp.fao.org/sd/vercon.pdf [Accessed 4 July 2016].

Food and Agriculture Organization (United Nations) (FAO), the International Fund for Agricultural Development (IFAD) and the World Food Programme (WFP). (2012). *The State of Food Insecurity in the World: Economic Growth is Necessary but not Sufficient to Accelerate Reduction of Hunger and Malnutrition.* Rome: FAO.

Food Crops Research Institute (FCRI). (2015. Human Resource Database- Lodged at FCRI. Unpublished.

Food Security and Nutrition Network and United States Agency International Development (USAID). (n.d). Data Quality Assurance. [Online]. Available at: http://www.fsnnetwork.org/sites/default/files/Data_quality_%20assurance_short.pdf [Accessed 3 June 2017].

Fossey, A. (2007). Bioethics in Agricultural Research and Research management. In G. Loebenstein and G. Thottappilly (ed). *Agricultural Research management*. Netherlands: Springer. Pp121-147.

Fry, J., Doiron, J., Létourneau, D, Perrier, L. and Perry, C. et al. (2017). Research Data Management Training Landscape in Canada: A White Paper. [Online]. Available at: https://portagenetwork.ca/wp-content/uploads/2017/02/Portage_ Training_White_Paper_EN.pdf [Accessed 21 March 2017].

Garrison D. R. (2000). Theoretical Challenges for Distance Education in the 21st Century: A Shift from Structural to Transactional Issues. In: *The International Review of Research in Open and Distance Learning,* 1 (1), 1-17.

Gay, L.R. and Airasian, P. (2003). *Educational Research:Competencies for Analysis and Application.* 7th (ed.). Upper Saddle River, NJ: Prentice Hall.

Genetic Resources Research Institute (GeRRI). (2015). Human Resource Database- Lodged at GeRRI. Unpublished.

Gibson, D. and Gross, J. (2013). Research Data Management in a Collaborative Network. Research Global. Issue 33, pp 12-15. [Online]. Available at: http://ro.ecu.edu.au/ecuworks2011/586 [Accessed 30 June 2016].

Golafshani, N. (2003). Understanding Reliability and validity in Qualitative research. *The Qualitative Report*, 8(4), 597-606.

Gold, A. (2007). Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and Data Challenges, Roles and Actions for Libraries. *D-Lib Magazine*, 13 (9/10).

Gorman, G.E. and Clayton, P. (2005). *Qualitative Research for the Information Professional: A Practical Handbook.* London: Facet Publication.

Government of Canada. (2016). Tri-Agency Statement of Principles on Digital Data Management. [Online]. Available at: http://www.science.gc.ca [Accessed 23 March 2017].

Gravetter, F,J, and Forzano, L.B. (2011). Research Methods for the Behavioural Sciences. [Online]. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.6907andrep=rep1andtype=pdf [Accessed 4 November 2016].

Grebmer, K.V. and Spielman, D.J. (2004). Public-Private Partnerships in Agriculture Research: An Analysis of Challenges Facing Industry and the Consultative Group on International Agricultural Research. Washington, D.C: International Food Policy Research Institute.

Greene, J.C. (2007). *Mixed Methods in Social Inquiry*. New York: Wiley.

Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. *Public Opinion Quarterly*, 72(2), 167-189.

Guba, E.G. (1990). The Alternative Paradigm Dialog. In E.G. Guba (ed.). *The Paradigm Dialog*, pp. 17-30. Newbury Park, CA: Sage.

Haas, J. and Murphy,S. (2009). E-Science and Libraries: Finding the Right Path. *Issues in Science and Technology Librarianship*, 57. Retrieved July 5, 2016 from http://www.istl.org/09-spring/viewpoint1.html.

Halbert, M. (2013). Prospects of Research Data Management. In *DataRe. Research Data Management: Principles, Practices, and Prospects.* Washington, DC: Council on Library and Information Resources. 1-15. [Online]. Available at: http://creativecommons.org/licenses/by-sa/3.0/ [Accessed 20 May 2016].

Harvey, R. (2006). Appraisal and Selection. In S. Ross and M. Day (Eds.). DCC Digital Curation Manual. Edinburgh, UK: Digital Curation Centre. [Online]. Available at: http://www.dcc.ac.uk/resource/curation-manual/chapters/appraisal-and-selection [Accessed 23 June 2016].

Harwell, M.R. (2011). Research Design in Qualitative/Quantitative/Mixed Methods. In: C.F. Clifton and R. C. Serlin (Eds.). *The SAGE Handbook for Research in Education: Pursuing Ideas as the Keystone of Exemplary Inquiry*, pp.147-163.Minnesota: Sage.

Heidorn, P.B. (2011). The Emerging Role of Libraries in Data Curation and E-Science. *Journal of Library Administration*, 51: (7-8), 662-672.

Henning, E. (2004). *Finding your way in Qualitative Research*. Pretoria: van Schalk Publishers.

Henty, M. (2014). Dreaming of Data: The Library's Role in Supporting E-Research and Data Management. [Online]. Available at: https://digitalcollections.anu.edu.au/bitstream/1885/47617/5/henty_alia_08.pdf [Accessed 3 January 2016].

Herren, R.V. and Donahue, R.L. (1991). *The Agriculture Dictionary*. New York: Delma.

Hey, T. and Trefethen, A. (2003). The Data Deluge: An e-Science Perspective. In F. Berman, G.C. Fox, and T. Hey, (Eds), *Grid Computing: Making the Global Infrastructure a reality*. New York: Wiley.

Higgins, S. (2008). The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, 3(1). [Online]. Available at: http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48 [Accessed 24 March 2016].

Higgins, S. (2011). Digital Curation: Emergence of a new Discipline. *The International Journal of Digital Curation*, 6(2), 78-88.

Higgins, S. (2012). The Lifecycle of Data Management. In G. Pryor. (Ed.). *Managing Research Data*. London: Facet Publishing.

Higman, R. and Pinfield, S. (2015). Research Data Management and Openness: The Role of Data Sharing in Developing Institutional Policies and Practices. *Program: Electronic Library and Information System,* 49 (4), 364-381.

Hinnant, C., Stvilia, B., Wu, S., Worrall, A., Burnett, K., Burnett, G., Kazmer, M. and Marty, P. (2012). Data Curation in Scientific Teams: an Exploratory Study of Condensed Matter Physics at a National Science Lab. In J. E. Mai (Chair), *Proceedings of iConference 2012* (pp.498-500). New York, NY: ACM.

Hodson, S. and Jones, S. (2013). Seven Rules of Successful Research Data Management in Universities. [Online]. Available at: https://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips [Accessed 19 September 2017].

Hoq, K.M.G. and Akter, R. (2012). Knowledge Management in Universities: Role of Knowledge Workers. *Bangladesh Journal of Library and Information Science*, 2, 92-102.

Hughes, C. (n.d). Qualitative and Quantitative Approaches. [Online]. Available at: http://dutmoodle.dut.ac.za/moodle/pluginfile.php/30889/mod_resource/content/0/Hughes_-_Qualitative_and_quantitative_approaches.pdf [Accessed 27 November 2016].

Humphrey, C. (2014). Are Libraries Organized to Provide Research Data Management Services? Preserving Research Data in Canada. [Online]. Available at: http://preservingresearchdataincanada.net/2014/12/04/are-libraries-organized-to-provide-research-data-management-services/ [Accessed 12 December 2015].

Indian Council of Agricultural Research (ICAR), (2014). ICAR Guidelines for Internal Evaluation and Forwarding Research Papers to Scientific Journals and Data Management in ICRAR Institutes. [Online]. Available at: http://www.icar.org.in/files/ICAR-Guidelines-Research-Papers-2014.pdf [Accessed 5 June 2017].

Industrial Research Institute. (2010). *Research Management*. Michigan: Industrial Research Institute.

Ingram, C. (2016). How and Why you should Manage your Research Data: A Guide for Researcher. [Online]. Available at: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data [Accessed 2 February 2016].

Interagency Working Group on Digital Data. (2009). Harnessing the Power of Digital Data for Science and Society Report of the Interagency Working Group. [Online]. Available at: https://www.nitrd.gov/About/Harnessing_Power_Web.pdf [Accessed 26 March 2017].

International Standards Organization (ISO). (2001). *ISO 15489-1:2001 Information and Documentation: Records Management Standard. Part 1: General*. Geneva: International Organization for Standardization.

Inter-university Consortium for Political and Social Research (ICPSR). (n.d). ICPSR Collection Development Policy. [Online]. Available at: http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html [Accessed 23 June 2016].

ISO 9001: 2008. (2008). Quality Management Systems – Requirements. [Online]. Available at: http://intra.itiltd-india.com/quality/ISOStandards/ISO9001-2008-STD.pdf [Accessed 13 June 2017].

Jahnke, L. M. and Asher, A. (2012). The Problem of Data: Data Management and Curation Practices among University researchers. In. L. Jahnke, A. Asher and S. Keralis. *The Problem of Data*. Washington, DC: Council on Library and Information Resources. [Online]. Available at: https://www.clir.org/pubs/reports/pub154/pub154.pdf [Accessed 3 September 2016].

Jahnke, L.M. and Asher, A. (2013). Dilemmas of Digital Stewardship: Research Ethics and the Problems of Data Sharing. In. *Research Data Management: Principles, Practices, and Prospects.* Washington, DC: Council of Library and Information Resources.

Jao, I.., Kombe, F., Mwalukore, S., Bull, S., Parker, M., Kamuya, D., Molyneux, S. and Marsh, V. (2015). Research Stakeholders' View on Benefits and Challenges for Public Health Research Data Sharing in Kenya: The Importance of Trust and Social Relations. [Online]. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135545 [Accessed 20 April 2016].

Jeffery, K. G. (2012). Current Research Information System (CRIS) in 2020. In: CRIS2012: *11th International Conference on Current Research Information Systems*, Prague, June 6-9, 2012. [Online]. Available at: http://dspacecris.eurocris.org/handle/11366/119 [Accessed 19 September 2017].

Jetten, M. (2014). LIBER Case Study: Research Data Management at Radboud University. [Online]. Available at: http://libereurope.eu/wp-content/uploads/2014/06/LIBER-Case-Study-Radboud.pdf [Accessed 26 April 2016].

Johanson, G. A. and Brooks, G.P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Education and Psychological measurement*, 70, 394-400.

Johnson, R. B., Onwuegbuzie, A. and Turner, L. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1 (2), 112-133.

Johnson, R.B. and Turner, L.A. (2003). Data Collection Strategies in Mixed Methods Research. In: A. Tashakkori and C. Teddlie (Eds.). *Handbook of Mixed Methods in Social and Behavioral research*, pp. 297-319. Thousand Oaks, CA: Sage.

Johnson, S. (2011). Importance of Data Analysis in Research. [Online]. Available at: http://dissertation-help-uk.blogspot.co.za/2011/12/imoptance of data analysis-in-research.html [Accessed 19 May 2017].

Johnsson, M. and Ahlfeldt, J. (2015). Research Libraries and Research Data Management within the Humanities and Social Sciences. [Online]. Available at: http://portal.research.lu.se/ws/files/6286782/5050466.pdf [Accessed 25 July 2017].

Joint, N. (2008). Current Research Information Sytems, Open Access Repositories and Libraries. *Library Review*, 57(8), 570-575.

Jones, S. (2013). Bringing it all Together: A Case Study on the Improvement of Research Data Management at Monach University. Data Curation Centre Research Data Management Services Case studies. Edinburgh. [Online]. Available at: http://www.dcc.ac.uk/resources/developing-rdm-services [Accessed 24 November 2016].

Jones, S., Guy, M. and Pickton, M. (2013). *Research Data Management for Librarians*. London: Data Curation Centre.

Jones, S., Pryor, G and Whyte, A. (2013). *How to Develop Research data management Services: A Guide for HEI's.* Edinburgh: Digital Curation Centre. [Online]. Available at: http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services [Accessed 23 September 2015].

Just, E. and Whitaker, S. (2016). The 3 Challenges of Translational and Clinical Research Data Management and a Strategy to Succeed. [Online]. Available at: https://www.healthcatalyst.com/addressing-challenges-clinical-research-data-management [Accessed 20 August 2016].

Kahn, M., Higgs, R. Davidson, J. and Jones, S. (2014). Research Data Management in South Africa: How we Shape Up. *Australian Academic and research Libraries*, 45(4), 296-308. [Online]. Available at: http://dx.doi.org/10.1080/00048623.2014.951910 [Accessed 30 March 2016].

Kalpesh, J. (2013a), Advantages and Disadvantages of Audio Recording Interviews. [Online]. Available at: http://www.dissertationindia.com/blog/advantages-and-disadvantages-of-audio-recording-interviews-2 [Accessed 6 October 2017].

Kalpesh, J. (2013b). Importance of Data Analysis in Research. [Online]. Available at: http://www.dissertationindia.com/blog/importance-of-data-analysis-in-research [Accessed 19 May 2017].

Kapahi, R. (2014). Difference between Census and Sampling. [Online]. Available at: http://www.slideshare.net/raghavkapahi/difference-between-census-and [Accessed 28 November 2016].

Karasti, H., Baker, K. and Hakola, E. (2006). Enriching the Notion of Data Curation in E-science: Data Managing and Information Infrastructure in the Long Term Ecological Research (LTER) Network. [Online]. Available at: http://www.worldpece.org/sites/default/files/57-227-1-PB.pdf [Accessed 3 May 2016].

Karick, A. (2014). Research Data Management: Guidelines and Planning for Researchers. [Online]. Available at: https://www.swinburne.edu.au/media/swinburneeduau/ research/docs/pdfs/Research_Data_Management_Guidelines.pdf [Accessed 4 October 2017].

Karolis, G. (2013). Understanding the Impact of Releasing and Re-using Open Government Data: European Public Sector Information Platform Topic Report No. 2013/08. [Online]. Available at: https://www.europeandataportal.eu/sites/default/files/2013_understanding_the_impact_of_releasing_and_re_using_open_data.pdf [Accessed 23 July 2016].

Kassam, A. and Odame, H.H. (2002). Listening to Stakeholders: Agricultural Research and Rural Radio Linkages. ISNAR briefing paper no. 48. Retrieved July7, 2016 from http://pdf.usaid.gov/pdf_docs/Pnacq553.pdf.

Kedemi, R. (2017). Overview of ‘A KARI and KAINet Institutional Repository Development: Case Study. [Online]. Available at: http://aims.fao.org/capacity-development/webinars/overview-kari-and-kainet-institutional-repository-development-case [Accessed 15 June 2017].

Kennan, M. A. (2016). Data Management: Knowledge and Skills Required in Research, Scientific and Technical Organizations. [Online]. Available at: http://library.ifla.org/1466/1/221-kennan-en.pdf [Accessed 27 June 2017].

Kenya Agricultural and Livestock Research Organization (KALRO). (2016). Kenya Agricultural and Livestock Research Organization-Home page. [Online]. Available at: http://www.kalro.org/ [Accessed 27 March 2016].

Kenya Agricultural and Livestock Research Organization (KALRO). (2016). *Service Charter (June 2016)*. Nairobi, Kenya: KALRO.

Kenya Agricultural and Livestock Research Organization (KALRO)-Tea Research Institute. (2016). *Information for Visitors*. Kericho, Kenya: Tea Research Institute.

Kenya Agricultural Research Institute (KARI). (2012). *Policy Responses to Food Crisis in Kenya*. USA, Washington: Food Security Portal.

Kenya, Republic of. (2002). *National Development Plan 2002-2008*. Nairobi: Government Printers.

Kenya, Republic of. (2007). Kenya Vision 2030. Nairobi: Government Printers. [Online]. Available at: http://www.vision2030.go.ke/cms/vds/Popular_Version.pdf [Accessed 2 May 2017].

Kenya, Republic of. (2012). National Agricultural Research System Policy. [Online]. Available at: http://extwprlegs1.fao.org/docs/pdf/ken147873.pdf [Accessed 13 November 2016].

Kenya, Republic of. (2013). Kenya Agricultural and Livestock Research Institute (KALRO) ACT NO 17 OF 2013. [Online]. Available at: https://www.stak.or.ke/wp-content/uploads/2016/08/The-Kenya-Agricultural-and-Livestock-Research-Act-2013.pdf [Accessed 13 June 2017].

Kenya Republic of: Agricultural Sector Coordinating Unit (ASCU). (2012). National Agricultural Research System Policy. [Online]. Available at: http://faolex.fao.org/docs/pdf/ken147873.pdf [Accessed 3 July 2016].

Kenya, Republic of: Ministry of Agriculture, Livestock and Fisheries. (2010). *Agricultural Sector Development Strategy 2010-2020*. Nairobi: Government Printers.

Keralis, S.D.C., Stark, S., Halbert, M. and Moen, W.E. (2013). Research Data Management in Policy and Practice: The DataRes Project. In Council on Library and Information Resources. *Research Data Management: Principles, Practices and Prospects*. Washing, DC: Council on Library and Information Resources.

Kerlinger, F.N. (1973). *Foundations of Behavioral Research*. 3rd (ed.). New York, NY: Holt, Rinehart and Winston.

Kerlinger, F.N. (1979). *Behavioral Research: A conceptual Approach*. 2nd (ed.). New York: Holt, Rinehart and Winston.

Keya, S.O., Makau, B.F., Mani, J. and Omari, I.M. (Eds.). (1989). *Guidelines for the Formulation of Research Project Proposal*. Nairobi: The National Council for Science and technology and The International Development Research Centre.

Kim, Y., Addom, B. and Stanton, J. (2011). Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure, *The International Journal of Digital Curation*, 6, 125-138.

King Abdullah Medical City. (2017). Registry and Data management Unit. [Online]. Available at: http://www.kamc.med.sa/index.php/en/departments/mass-gathering-epidemiology-and-statistics-department/registry-and-data-management-unit [Accessed 8 June 2017].

King, G., Keohane, R.O. and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton, NJ: Princeton University Press.

King, N. (2004). Using Interviews in Qualitative Research. In C. Casell and G. Symon (Eds.). *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage, pp. 11-22.

Kiplang'at, J. (2004). *Diffusion of Information and Communication Technologies in Communication of Agricultural Information among Agricultural Researchers and Extension Workers in Kenya*. Unpublished D. Phil. Thesis. KwaZulu-Natal, SA: University of Zululand, Department of Library and Information Science.

Kirklees Council. (n.d). Research and Consultation Guidelines. [Online]. Available at: https://www.kirklees.gov.uk/community/yoursay/questionnaires.pdf [Accessed 28 November 2016].

Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data. (2016). Funding Research Data Management and related infrastructures: Knowledge Exchange and Science Europe Briefing Paper. [Online]. Available at: http://www.scienceeurope.org/wp-content/uploads/2016/05/SE-KE_Briefing_Paper_Funding_RDM.pdf [Accessed 12 August 2016].

Koltay, T. and University, S. I. (2015). Data Literacy in Support of Research. [Online]. Available at: https://libraryconnect.elsevier.com/articles/data-literacy-support-research-20 [Accessed 13 October 2017].

Koopman, M.M. and De Jager, K. (2016). Archiving South African Digital Research Data: How Ready are We? *South African Journal of Science*, 112 (7/8), 1-7.

Kothari, C.R. (2004). *Research Methodology: Methods and Techniques*. 2nd (ed.). New-Delhi: New age International Publishers.

Krishnaswami, O.R. and Ranganathan, M. (2010). *Methodology of Research in Social Science.* Mumbai: Himalaya Publishing House.

Kruse, F. and Thestrup, J.B. (2014). Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark. *Liber Quarterly*, 23(4), 310-335.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. 2nd (ed.). Chicago: University of Chicago Press.

Kumar, R. (2011). *Research Methodology: A Step-by-Step Guide for Beginners*. 3nd ed. London: Sage publications Ltd.

Kuusniemi, M. E., Heino, T. and Larmo, K. (n.d). How to Get Started with Research Data Management Training Services for the Academic Library? [Online]. Available at: http://www.iss.it/binary/eahi/cont/76_Mari_Elisa_Kuusniemi_Full_text.pdf [Accessed 16 April 2016].

Kvale, S. (1996). *Interviews.* London: Sage.

Lairumbi, G.M., Parker, M., Fitzpatrick, R. and Mike, E.M. (2011). Stakeholders Understanding of the Concepts of Benefit Sharing in Health Research in Kenya: A Qualitative Study. [Online]. Available at: https://bmcmedethics.biomedcentral.com/articles/10.1186/1472-6939-12-20 [Accessed 24 March 2017].

Laughton, P. and Du Plessis, T. (2013). Data Curation in the World Data System: Proposed Framework. *In Data Science Journal*, 12, 56-70.

Leedy, P.D. (1997). *Practical Research: Planning and Design*. 6th (ed.). New Jersey: Prentice-Hall, Inc.

Leedy, P.D. and Ormrod, J.E. (2005). *Practical Research: Planning and Design*. 8th (ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Leigh, G. (2014). Stuff and Data: Challenges for RDM in the Visual Arts. [Online]. Available at: http://research.uca.ac.uk/1739/ [Accessed 29 June 2017].

Lewis, M.J. (2010). Libraries and the Management of Research Data. In: *Envisioning Future Academic Library Services*. Facet Publishing: London. 145-168.

Lincoln, Y.S and Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lindquist, A. (2013). Advantages of Electronic Data Capture (EDC) over other Methods for Data Capture. [Online]. Available at: https://forteresearch.com/news/advantages-of-edc-over-other-methods-for-data-capture/ [Accessed 4 July 2017].

LoBiondo-Wood, G. and Haber, J. (2014). *Nursing Research: Methods and Critical Approach for Evidence based Practice*. St. Louis: Elsevier.

Loebenstein and Thottappilly. (2007). The Mission of Agricultural Research. [Online]. Available at: https://link.springer.com/chapter/10.1007%2F978-1-4020-6057-1_1#page-1 [Accessed 29 March 2017].

Lord, P. and Macdonald, A. (2003). E-science Curation Report: Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision. Prepared for the JISC Committee for the Support of Research. Twickenham, England: The Digital Archiving Consultancy Limited. [Online]. Available at: http://www.jisc.ac.uk/uploaded documents/e-scienceReportFinal.pdf [Accessed 2 September 2015].

Lord, P., Macdonald, A., Lyon, L. and Giaretta, D. (2004). From Data Deluge to Data Curation. In Proceedings of the 3rd UK e-Science. [Online]. Available at: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf [Accessed 8 August 2016].

Lötter, L. (2014). *Reflections on the RDM Position in South Africa: Development Research Data Management Services*. Cape Town, South Africa: Research Methodology Centre.

Loughborough University. (n.d). Data Presentation and Analysis. [Online]. Available at: http://wedc.lboro.ac.uk/resources/units/IRP_Unit%207_Data%20presentation_analysis.pdf [Accessed 12 June 2017].

Loughborough University. (2017). Scholarly Communications: Research Data Management. [Online]. Available at: http://www.lboro.ac.uk/research/scholcomms/researchdata/ [Accessed 24 September 2017].

Lynch, C.A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. [Online]. Available at: http://www.arl.org/resources/pubs/br/br226/br226lr.shtml [Accessed 11 August 2016].

Lynch, C. A. (2014). The Need for Research Data Inventories and the Vision for SHARE. *Information Standards Quarterly,* Summer 2014.

Lyon, L. (2012). The Informatics Transform: Re: Engineering Libraries for the Data Decade. *International Journal of Digital Curation*, 7(1), 126-138.

Lyon, L., Ball, A., Duke, M. and Day, M. (2012). Community Capability Model Framework. [Online]. Available at: http://communitymodel.sharepoint.com [Accessed 12 March 2016].

Lyon, L., Patel, M. and Takeda, K. (2014). Assessing Requirements for Research Data Management Support in Academic Libraries: Introducing a New Multi-Faceted Capability Tool. [Online]. Available at: http://ozk.unizd.hr/proceedings/ index.php/lida/article/view/121 [Accessed 28 November 2015].

Macanda, M., Rammulloa, M., and Bezuidenhout, R. (2015). Research Data Management at University of South Africa (UNISA). [Online]. Available at: http://hdl.handle.net/10500/139007 [Accessed 24 March 2017].

Macdonald, S. and Martinez-Uribe, L. (2010). A New Role for Academic Librarians: Data Curation. *EI Profesional de la Informacio'n*, 17(3): 273-280.

MacKenzie, S. (2014). Data Governance: Where Technology and Policy collide. In J. M. Ray (Ed.). *Research Data Management: Practical Strategies for Professionals*. Indiana: Purdue University Press.

Manda, P.A. (2002). Information and agricultural Development in Tanzania: a Critique. *Information Development,* 18(3): 181-188.

Marchionini, G., Lee, C. A. and Bowden, H. (2012). Curating for Quality: Ensuring Data Quality to Enable New Science. [Online]. Available at: https://ils.unc.edu/callee/curating-for-quality.pdf [Accessed 5 July 2017].

Margolis, J. (2003). Pragmatism's advantage. *Ars Disputandi*, 3(1), 300-326.

Maru, A. (2004). ICT/ICM in Agricultural Research and Development: Status in Sub-Saharan Africa. [Online]. Available at: http://www.fao.org/docs/eims/upload/210126/ICT_in_ARD_Regional_Status_2004_FARA.pdf [Accessed 3 June 2016].

Maseh, E.J. (2015). *Records Management Readiness for open Government in the Kenyan Judiciary*. Unpublished D. Phil. Thesis. KwaZulu-Natal, SA: University of KwaZulu-Natal, Pietermaritzburg, School of Social Sciences/Department of Information Studies.

Mason, M. and Culnan, M. (1995). *Ethics of Information Management*. New Delhi: Sage Publications.

Mathipa, E.R. (2015). Review of Pertinent Literature in Research. In E. R. Mathipa and M. T. Gumbo (Eds.). *Addressing Research Challenges: Making Headway for Developing Researchers*. Noordwyk, South Africa: Mosala-MASEDI Publishers and Booksellers.

Maxwell, J.A. (2013). *Qualitative Research Resign: An Interactive Approach*. (3rd ed.). Thousand Oaks, CA: Sage.

May, T. (1993). *Social Research: Issues, Methods and Process*. Buckingham: Open University Press.

McGeever, M., Whyte, A. and Molloy, L. (2015). Five Things you need to Know about Research Data Management and the Law. DCC Checklist on Legal Aspects of RDM. [Online]. Available at: http://www.dcc.ac.uk/resources/how-guides/rdm-law [Accessed 1 June 2017].

McLaughlin, E. (2016). Data Collection. [Online]. Available at: http://searchcio.techtarget.com/definition/data-collection [Accessed 28 November 2016].

McLaughlin, J.E., Bush, A.A., and Zeeman, J.M. (2016). Mixed Methods: Expanding Research Methodologies in Pharmacy Education. [Online]. Available at: http://dx.doi.org/10.1016/j.cptl.2016.06.015 [Accessed 16 June 2017].

Mcnamara, K. (2009). Improving Agricultural Productivity and Markets: The role of Information and Communication Technologies. Retrieved August 3, 2013 from http://siteresources.worldbank.org/INTARD/Resources/47WB_ARD_Role_ITC_Note47_web_0509v2.pdf.

Milovic, B. and Radojevic, V. (2015). Application of Data Mining in Agriculture. *Bulgarian Journal of Agricultural Science*, 21 (1), 26-34.

Miruka, M.K., Okello, J.J., Kirigua, V.O. and Murithi, F.M. (2015). The Role of the Kenya Agricultural Research Institute (KARI) in Attainment of Household Food Security in Kenya: A policy and Organization Review. [Online]. Available at: https://www.researchgate.net/profile/Maureen_Miruka/publication/257788782_The_role_of_the_Kenya_Agricultural_Research_Institute_KARI_in_the_attainment_of_household_food_security_in_Kenya_A_policy_and_organizational_review/links/558c685c08ae40781c204d78.pdf [Accessed 20 March 2017].

Mishler, E.G. (2000). Validation in Inquiry-Guided Research: The Role of exemplars in Narrative Studies. In B.M. Brizuela, J.P. Stewart, R.G. Carrillo, and J.G. Berger (Eds.). *Acts of Inquiry in Qualitative Research*, pp. 119-146. Cambridge, MA: Harvard Educational Review.

Molloy, L. and Snow, K. (2012). The Data Management Skills Support Initiatives: Synthesizing Postgraduate Training in Research Data Management. *International Journal of Digital Curation*, 7(2), 101-109.

Morgan, D. (1998). Practical Strategies for Combining Qualitative and Quantitative Methods: Applications for Health Research. *Qual, Health Research*, 8: 362-376.

Morgan, D. L. (2006). Connected Contributions as a Motivation for Combining Qualitative and Quantitative Methods. In: L. Curry, R. Shield and T. Wetle (Eds*). Improving Aging and Public Health research: Qualitative and Mixed methods.* Washing, DC: American Public Health Association and Gerontological Society of America, pp 53-63.

Morris, S. (2013). Issues in the Appraisal and Selection of Geospatial Data: A National Digital Stewardship Alliance (NDSA) Report. [Online]. Available at: http://www.digitalpreservation.gov/documents/NDSA_AppraisalSelection_report_final102413.pdf [Accessed 5 October 2016].

Morse, J., Wolfe, R. and Niehaus, L. (2006). Principles and Procedures of Maintaining Validity for Mixed-Method Design. In: L. Curry, R. Shield and T. Wetle. (Eds*). Improving Aging and Public Health research: Qualitative and Mixed methods*. Washing, DC: American Public Health Association and Gerontological Society of America, pp 65-78.

Moseti, I. M. (2015). *Strategies for Managing Scholarly Content at Universities in Kenya.* Unpublished D. Phil. Thesis. KwaZulu-Natal, SA: University of KwaZulu-Natal, Pietermaritzburg, School of Social Sciences/Department of Information Studies.

Mossink, W., Bijsterbosch, M. and Nortier, J. (2013). European Landscape Study of Research Data Management: Support Infrastructure Models for Data Management. [Online]. Available at: http://www.clarin.nl/sites/default/files/SIM4RDM% 20landscape% 20report_1.pdf [Accessed 15 March 2016].

Mugabe, J. (2001). Closing the Genetic and Digital Divides: Towards Africa's Common Agenda. [Online]. Available at: http://www.scienceafrica.co.za/2001/august/agenda.htm [Accessed 3 October 2016].

Mugata, R. (2014). KAINeT Initiative Opening Up Access to Agricultural Research Information in Kenya. [Online]. Available at: http://www.ciard.info/news-and-events/blog/kainet-initiative-opening-access-agricultural-research-information-kenya [Accessed 22 April 2016].

Mugenda, O.M. and Mugenda, A.G. (1999). *Research Methods: Quantitative and Qualitative Approaches.* Nairobi: African Centre for Technology Studies (ACTS) Press.

Mugenda, O.M. and Mugenda, A.G. (2003). *Research Methods: Quantitative and Qualitative Approaches.* Nairobi: African Centre for Technology Studies.

Mugwisi, T. (2013). The Information Needs and Challenges of Agricultural Researchers and Extension Workers in Zimbabwe. Unpublished D.Phil. Thesis. South Africa: University of Zululand, Department of Library and Information Science.

Muinde, F.N.N. and Gorman, G.E. (2009). Barriers to Open Access to Scientific Information in Kenya with Particular Reference to Agricultural Information. [Online]. Available at: http://www.ifla.org/annual-conference/ifla75/index.htm [Accessed 26 August 2016].

Mukiibi, E. (2016). Research Data Management and Research Data Re-use in Uganda. In 22nd Standing Conference of Eastern, Central and Southern Africa Library and Information Associations (SCECSAL XXII). Proceedings: *Digital Transformation and the Changing Role of Libraries and Information Centres in the Sustainable Development of Africa*. Hosted in Swaziland by Swaziland Library and Information Association 25-29 April 2016.

Mullins, J.L. (2014). The Policy and Institutional Framework. In J. M. Ray. (Ed.). *Research Data Management: Practical Strategies for Professionals*. Indiana: Purdue University Press.

Munyua, H. (2000). Application of Information Communication Technologies in the Agricultural Sector in Africa: A Gender Perspective. In E. Rathgeber and E. O. Adera. (Eds). *Gender and Information Revolution in Africa,* IDRC/ECA. pp. 85-123.

Murphy, J.P. (1990). *Pragmatism: From Peirce Davidson. Boulder*, CO: Westview.

Mutula, S. (2014). Impact and Evaluation of Research Data Management. [Online]. Available at: http://uir.unisa.ac.za/bitstream/handle/10500/13912/mutulastephen_impact%20....pdf?sequence=1andisAllowed=y [Accessed 5 July 2016].

National Archives and Records Administration (NARA). (2007). Strategic Directions: Appraisal Policy. [Online]. Available at: http://www.archives.gov/records-mgmt/initiatives/appraisal.html [Accessed 23 June 2016].

National Digital Stewardship Alliance (NDSA). (2003). Issues in the Appraisal and Selection of Geospatial Data. [Online]. Available at: http://www.digitalpreservation.-gov/documents/NDSA_Appraisal Selection_report_final102413.pdf [Accessed 14 January 2017].

National Endowment for the Humanities. (2013). Data Management Plans for NEH Office of Digital Humanities Proposals and Awards. [Online]. Available at https://www.clir.org/pubs/reports/pub160/pub160.pdf [Accessed 15 March 2017].

National Institutes of Health (NIH). (2003a). Final NIH Statement on Sharing Research Data. [Online]. Available at: http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html [Accessed 30 March 2017].

National Institutes of Health (NIH). (2003b). NIH Data Sharing Policy and Implementation Guidance. [Online]. Available at: http://grants.nih.gov/grants/policy/data_sharing/data_guidance.htm [Accessed 3 June 2017].

National Institutes of Health (NIH). (2008). Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research. [Online]. Available at: http://www.ucop.edu/raohome/cgmemos/08-05.pdf [Accessed 7 May 2016].

National Library of Australia. (2003). Guidelines for the Preservation of Digital Information. [Online]. Available at: http://unesdoc.unesco.org/images/0013/001300/130071e.pdf [Accessed 30 June 2016].

National Research Foundation (NRF). (2015). Databases. [Online]. Available at: http://www.nrf.ac.za/information-resources/databases [Accessed 15 February 2017].

National Science Foundation (NSF). (2007a). Cyberinfrastructure Vision for 21st Century Discovery. [Online]. Available at: http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf [Accessed 6 July 2016].

National Science Foundation (NSF). (2007b). Sustainable Digital Data Preservation and Access Network Partners (DataNet): Program Solicitation NSF 07-601. National Science Foundation, Office of Cyberinfrastructure. [Online]. Available at: http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm [Accessed 26 May 2016].

National Science Foundation (NSF). (2011). Digital Research Data Sharing and Management, No.NSB-11-79. [Online]. Available at: https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf [Accessed 20 July 2016].

National University of Ireland Galway. (n.d). Backup. [Online]. Available at: https://www.nuigalway.ie/information-solutions-services/servicesforstaff/desktopsupport/faq/backup/ [Accessed 27 July 2017].

National University of Singapore. (2017). Research Data Curation and Management: Data Management Policy. [Online]. Available at: http://libguides.nus.edu.sg /c.php?g=338056andp=2325062 [Accessed 30 June 2017].

Ndemo, B. (2016). Wheat should be Central to Kenya's food Security Strategy. The Daily Nation (digital). [Online]. Available at: http://www.nation.co.ke/oped/blogs/dot9/ ndemo/-/2274486/2986916/-/bporh/-/index.html [Accessed 25 April 2016].

Neuman, W.L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. 4th (ed.). Boston: Ally and Beacon.

Newton, M.P., Miller, C.C. and Bracke, M.S. (2011). Librarian role in Institutional Repositories Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Management*, 36 (1), 53-67.

Niu, J. (2014). Appraisal and Selection for Digital Curation. *International Journal of Digital Curation*, 9(2), 65-82.

Njuguna, F.W and Itegi, F.M. (2013). Research in Institutions of Higher Education in Africa: Challenges and Prospects. *European Scientific Journal*, December 2013/special edition Vol.1.

North Carolina State University (NCSU) Libraries. (n.d). Defining Research Data: Data Management. [Online]. Available at: http://www.lib.ncsu.edu [Accessed 23 March 2017].

Northey, M., Tepperman, L. and Albanese, P. (2012). *Making Sense: Social Sciences*. 5th (ed.). Ontario, Canada: Oxford University Press.

O'Leary, Z. (2004). *The Essential Guide to Doing Research*. London: Sage.

Olsen, M.E., Lodwick, D.G, and Dunlap, R.E. (1992). *Viewing the World Ecologically*. Boulder, CO: Westview Press.

Olum, G. H. (2013). *Enhancing Migration Data management in Kenya: Assessment and Recommendations.* Nairobi: ACP Observatory on Migration.

Onwuegbuzie, A.J., Collins, K.M.T., Leech, N.L., Dellinger, A.B. and Jialo, Q.G. (2010). *A Meta-framework for Conducting Mixed Research Synthesis for Stress and Coping Researcher and beyond.* London: Information Publ.

Organisation for Economic Co-operation and Development (OECD). (2004). OECD Principles of Corporate Governance. [Online]. Available at: https://www.oecd.org/corporate/ca/corporategovernanceprinciples/31557724.pdf [Accessed 7 May 2017].

Organisation for Economic Co-operation and Development (OECD). (2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. [Online]. Available at: http://www.oecd.org/sti/sci-tech/38500813.pdf [Accessed 4 July 2016].

Outsource2india. (2016). Different Types of Questions in Questionnaire Design. [Online]. Available at: https://www.outsource2india.com/kpo/articles/questionnaire-types-of-questions.asp [Accessed 28 November 2016].

Palathingal, L., Dascalu, S., Harris, F.C. and Varol, Y. (2015). A Brief of Data Curation Literature. [Online]. Available at: https://www.cse.unr.edu/~fredh/papers/conf/143-absodcl/paper.pdf [Accessed 20 May 2016].

Palmer, C.L., Cragin, M.H., Heidorn, P.B. and Smith, L.C. (2007). Data Curation for the Long Tail of Science: The Case of Environment Studies. [Online]. Available at: http://www.researchgate.net/publication/228656107 [Accessed 28 March 2016].

Pandav, R., Mehta, A., Belle, S. H., Martin, D. E., Chandra, V., et al. (2002). Data Management and Quality Assurance for an International Project: The Indo-US Cross-National Dementia Epidemiology Study. [Online]. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.532.4648andrep=rep1andtype=pdf [Accessed 31 July 2017].

Parker, R. (2012). What the Library Did Next: Strengthening Our Visibility in Research Support. [Online]. Available at: http://www.vala.org.au/direct-download/vala2012-proceedings/435-vala2012-session-1-parker-paper/file [Accessed 18 July 2016].

Patel, S. (2015). The research Paradigm-Methodology, Epistemology and Ontology-Explained in Simple Language. [Online]. Available at: http://salmapatel.co.uk/academia/the-research-paradigm-methodology-epistemology-and-ontology-explained-in-simple-language [Accessed 15 November 2016].

Patel, D. (2016). Research Data Management: A Conceptual Framework. *Library Review*, 65 (4/5), 226-241. [Online]. Available at: http://www.emeraldinsight.com/doi/pdfplus/10.1108/LR-01-2016-0001 [Accessed 11 October 2017].

Patrick, M., Byatt, D., De Luca, F., White, W. Rumsey, S. and Wilson, A. J. (2013). Research Data Management Training for Support Staff. A DaMaRO Project Survey. [Online]. Available at: http://damaro.oucs.ox.ac.uk/docs/Research_Data_Management Training _for%20_Support_Staff_-_a_DaMaRO_Project_Survey.pdf [Accessed 5 July 2016].

Patton, M.Q. (1990). Qualitative Evaluation and Research Methods. 2nd (ed.). Newbury Park, CA: Sage.

Patton, M.Q. (2002). *Qualitative Evaluation and Research Methods*. 3rd (ed.). Thousand Oaks, CA: Sage Publications Inc.

Pautasso, M. (2013). Ten Simple Rules for Writing a Literature Review. *Plos Comput Biol*, 9(7), 34-47.

Pauw, K. and Thurlow, J. (2011). *The Role of Agricultural Growth in Reducing Poverty and Hunger: The Case of Tanzania.* International Food Policy Research Institute (IFPRI) 2020 Conference Brief No. 21. Washington DC. IFPRI.

Pennock, M. (2007). Digital Curation: A Life-cycle Approach to Managing and Preserving Usable Digital Information. [Online]. Available at: http://www.dcc.ac.uk h ttp://www.ukoln.ac.uk/ [Accessed 30 June 2016].

Peters, C. and Dryden, A.R. (2011). Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston. *Science and Technology Libraries,* 30(4), 387- 403.

Phelan, C. and Wren, J. (n.d). Exploring Reliability in Academic Assessment. [Online]. Available at: https://www.uni.edu/chfasoa/reliabilityandvalidity.htm [Accessed 29 November 2016].

Pickard, A.J. (2007). *Research Methods in Information*. London: Facet Publishing.

Pienaar, H. (2010). Survey of Research Data Management Practices at the University of Pretoria, South Africa: October 2009-March 2010. [Online]. Available at: http://www.ais.up.ac.za/profile/heila_pienaar/index.htm [Accessed 28 March 2016].

Pinfield, S., Cox, A.M. and Smith, J. (2014). Research Data Management and Libraries: Relationship, Activities, Drivers and Influences, *PLoS ONE, 9(12)*. [Online]. Available at: http://eprints.whiterose.ac.uk/82499/1/PLOS%20ONE% 20final%20published% 20version.pdf [Accessed 25 March 2016].

Polit, D. F. and Beck, C. T. (2001). *Essentials of Nursing Research: Methods, Appraisal and Utilization.* 5th (ed.). Pliladelphia: Lippincott.

Polit, D.F. and Beck, C.T. (2004). *Nursing Research: Principles and Methods*. 7th (ed.). Pliladelphia: Lippincott.

Princeton University. (2017). Research Data Security. [Online]. Available at: https://www.princeton.edu/ria/human-research-protection/data/ [Accessed 11 October 2017].

Pryor, G. (2012). Why Manage Research Data? In G. Pryor. (Ed.). *Managing Research Data*. London: Facet Publishing.

Qin, J. (2013). Infrastructure, Standards, and Policies for Research Data Management. [Online]. Available at: http://surface.syr.edu/cgi/viewcontent.cgi?article=1164andcontext=istpub [Accessed 16 April 2016].

Queensborough Community College. (2016). Academic Works: Your Institutional Repository: Benefits of Digital Repositories. [Online]. Available at: http://qcc.libguides.com/c.php?g=301716andp=2019401 [Accessed 4 July 2016].

Rajasekar, S., Philominathan, P. and Chinnathambi, V. (2006). Research Methodology. [Online]. Available at: https://arxiv.org/pdf/physics/0601009.pdf [Accessed 14 November 2016].

Ray, J. M. (2014). Introduction to Research Data Management. In. J. M. Ray. (Ed). *Research Data Management: Practical Strategies for Information Professionals.* West Lafayette, Indiana: Purdue University Press.

RECODE Project Consortium. (2014). Policy Recommendations for Open Access to Research Data. [Online]. Available at: http://www.recodeproject.eu [Accessed 3 August 2016].

Reisch, J. and Webster, J. (2004). Quality Control Measures. [Online]. Available at: http://www.alz.washington.edu/NONMEMBER/FALL04/texas.pdf [Accessed 31 July 2017].

Resnik, D.B. (2015). What is Ethics in Research and Why is it Important? [Online]. Available at: [Online]. Available at: http://www.niehs.nih.gov/research/resources/bioethics/whatis/ [Accessed 30 November 2016].

Riley, J. (2009). Seeing Standards: a Visualization of the Metadata Universe. [Online]. Available at: http://scimaps.org/mapdetail/seeing_standards_a_v_130 [Accessed 3 July 2016].

Rossman, G.B. and Rollis, S.F. (2003). *Learning in the Field: An Introduction to Qualitative Research*. 2nd (ed.). Thousand Oaks, CA: Sage.

Rossman, G.B. and Wilson, B.L. (1985). Numbers and Words: Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study. *Evaluation Review*, 9(5), 627-643.

Sandler, C. (2013). Agricultural Terms: Part I. [Online]. Available at: http://blog.machinefinder.com/11045/agricultural-terms-part-i [Accessed 25 August 2016].

Saunders, M., Lewis, P. and Thornhill, A. (2012). *Research methods for Business Students.* 6th (ed.). Edinburgh Gate, England: Pearson Education Limited.

Schellenberg, T.R. (1956). *The Appraisal of Modern Public Records (Bulletins of the National Archives No. 8).* Washington, DC: Government Printing Office.

Schmidt, B. and Shearer, K. (2016). Librarians' Competencies Profile for Research Data Management. [Online]. Available at: https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf [Accessed 26 May 2017].

Scholarly Publishing and Academic Resource Coalition. (n.d). Institutional Repositories and Open Access-The Future of Scholarly Communication. [Online]. Available at: http://www.budapestopenaccessinitiative.org/pdf/David_Prosser.pdf [Accessed 4 October 2016].

Searle, S. (2011). University Libraries and Research Data Management: Developing Knowledge, Skill and Careers. [Online]. Available at: https://www.monash.edu/__data/assets/pdf_file/0007/157786/caul-pres-final_nb.pdf [Accessed 30 June 2017].

Sekaran, U. (2003). *Research Methods for Business: A Skill Building Approach*. (4th ed.). New York: John Wiley and Sons, Inc.

Sewerin, C. (2015). Research Data Management Faculty Practices: A Canadian Perspective. Proceedings of the IATUL Conference. Paper 2. [Online]. Available at: http://docs.lib.purdue.edu/iatul/2015/mvd/2 [Accessed 21 March 2017].

Shakeri, S. (2013). Data Curation Perspectives and Practices of Researchers at Kent State University's Liquid Crystal Institute: A case Study (Masters Dissertation). [Online]. Available at: http://lod-lam.slis.kent.edu/documents/StudentResearch/ShadiShakeri/Shakeri_Thesis_Corrected_Nov26_PDF-A.pdf [Accessed 28 April 2016].

Sherbak, T. (2014). 6 Data Management Challenges Facing IT. [Online]. Available at: http://blog.datagravity.com/6-data-management-challenges-must-overcome/ [Accessed 19 August 2016].

Smith, M. L. (1991). Meanings of Test Preparation. *American Educational Research journal*, 28 (3), 521-542.

Smith, M. L. (2014). Data Governance: Where Technology and Policy Collide. In Ray, J.M. (ed). *Research Data Management: Practical Strategies for Information Professionals*. Indiana, US: Purdue University.

Social Sciences and Humanities Research Council (SSHRC). (2015). Tri-Agencies Adopt Statement of Principles on Digital Data Management. [Online]. Available at: http://www.sshrc-crsh.gc.ca [Accessed 23 March 2017].

The Society of College, National and University Libraries (SCONUL). (2015). Research Data Management: Briefing for Library Directors. [Online]. Available at: http://www.sconul.ac.uk/sites/default/files/documents/SCONUL%20RDM%20briefing.pdf [Accessed 24 July 2016].

Sörlin, S. and Vessuri, H. (2007). Introduction: The Democratic Deficit of Knowledge Economies. In: S. Sörlin, and H. Vessuri (Eds.). *Knowledge Society vs. Knowledge Economy: Knowledge, Power, and Politics.* New York: Palgrave Macmillan.

Stake, R.E. (1995). *The Art of Case Study Analysis.* New York: Guilford Press.

Stanford Libraries. (n.d(a)). Case Study: Data Storage and Backup. [Online]. Available at: https://library.stanford.edu/research/data-management-services/case-studies/case-study-data-storage-and-backup [Accessed 27 June 2017].

Stanford Libraries. (n.d(b)). Data Management Plans. [Online]. Available at: https://library.stanford.edu/research/data-management-services/data-management-plans [Accessed 11 October 2017].

Statistics Canada. (2015). Data Collection Capture and Coding. [Online]. Available at: http://www.statcan.gc.ca/pub/12-539-x/2009001/collection-collecte-eng.htm [Accessed 30 June 2017].

Steinhart, G. (2011). DataStaR: A Data Sharing and Publication Infrastructure to Support Research. *An International Journal for Information Specialists in Agriculture, Natural Resources and the environment*, 4(1), 16-20.

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D. and Kramer, S. (2012). Prepared to Plan? A Snapshort of Researcher Readiness to Address Data Management Planning Requirements. *Journal of eScience Librarianship*, 1(2).

Stenbacka, C. (2001). Qualitative Research Requires Quality Concepts of its Own. *Management Decision*, 39 (7), 551-555.

Stienen, J., Bruinsma, W. and Neuman, F. (2007). How ICT can make a Difference in Agricultural Livelihoods. [Online]. Available at: http://www.esteemconsults.com/downloads/ICT%20and%20agricultural%20 Development.pdf [Accessed 5 July 2016].

Streiner, D.L. (2003). Being Inconsistent about Consistency: When Coefficient Alpha Does and Doesn't Matter. *Journal of Personality Assessment*, 80(3), 217-222.

Swan, A. and Brown, S. (2008). The Skills, Role and Career Structure of Data Scientists and Curators: Assessment of Current Practice and Future needs: Key Perspectives. [Online]. Available at: http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/ dataskillscareersfinalreport.pdf [Accessed 29 July 2017].

Swanson, R.A. (2013). *Theory Building in Applied Disciplines*. San Francisco: Berrett-Koehler Publishers.

Tashakkori, A. and Teddlie, C. (2003). Handbook *off Mixed Methods in Social and Behavioral Research.* Thousand Oaks: Sage.

Tashakkori, A. and Teddlie, C. (2010). Overview of Contemporary Issues in Mixed Methods Research. In: A. Tashakkori and C. Teddlie (Eds.). *Handbook of mixed Methods in Social and Behavioral Research*, pp. 1-41. Thousand Oaks, CA: Sage.

Taylor, R. (2014). The SKA and the Rise of Big Data in Radio Astronomy. Presentation given at the Research Africa Conference 2014, Cape Town, South Africa. [Online]. Available at: http//www.eresearch.ac.za/presentations-2014-day-three [Accessed 26 September 2016].

Tea Research Institute (TRI). (2015). Human Resource Database- Lodged at TRI. Unpublished

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Read, E. et al. (2011). Data Sharing by Scientists: Practices and Perceptions. PLoS One 6(6). [Online]. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101 [Accessed 20 November 2016].

Tenopir, C., Birch, B., and Allard, S. (2012). Academic Libraries and Research Data Services: Practices and Plans for the Future. *Association of College and Research Libraries*. [Online]. Available at: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/ publications/whitepapers/Tenopir_Birch_Allard.pdf [Accessed 1 April 2016].

Tenopir, C., Sandusky, R. J., Allard, S. and Birch, B. (2014). Research Data Management Services in Academic Research Libraries and Perceptions of Librarians. *Library and Information Science Research*, 36(2), 84-90.

Thanos, C. (2010). Global Research Data Infrastructures: The GRDI 2010 Vision. [Online]. Available at: http://www.grdi2020.eu/Repository/FileScaricati/fc14b1f7-b8a3-41f8-9e1e-fd803d28ba76.pdf [Accessed 12 September 2016].

Tibbo, H.R. (2003). On the nature and importance of archiving in the digital age. In M. Zelkowitz (ed.). *Advances in Computers: Vol. 57. Information Repositories.* Retrieved June 23, 2016 from doi: 10.1016/S0065-2458(03)57001-2.

Tjalsma, H. and Rombouts, J. (2011). Selection of Research Data: Guidelines for Appraising and Selecting Research data. [Online]. Available at: file:///C:/Users/216056748/Downloads/DANSselectionofresearchdata% 20(1).pdf [Accessed 2 June 2017].

Treloar, A. and Wilkinson, R. (2008). Rethinking Metadata Creation and Management in a Data-Driven Research World, *Proceedings of the Fourth IEEE International Conference on eScience held on 10-12 December 2008,* organized by the Institute for Electrical and Electronics Engineers. [Online]. Available at: http://doi.ieeecomputersociety.org/10.1109/eScience.2008.41 [Accessed 3 July 2016].

Trochim, W.M.K. (2006). Research Methods Knowledge Base: Qualitative Measures. [Online]. Available at: http://www.socialresearchmethods.net/kb/qual.php [Accessed 30 November 2016].

UK Data Archive. (2017). Create and Manage Data: Quality Assurance. [Online]. Available at: http://www.data-archive.ac.uk/create-manage/format/quality [Accessed 1 August 2017].

UK National Archives. (2012). Appraisal Policy. [Online]. Available at: http://www.nationalarchives.gov.uk/documents/information-management/appraisal_policy.pdf [Accessed 23 June 2016].

United Nations Development Programme (UNDP). (2012). The State of Food and Agriculture: Investing in Agriculture for a Better Future. [Online]. Available at: http://www.fao.org/docrep/017/i3028e/i3028e.pdf [Accessed 14 January 2017].

United Nations Environmental Programme (UNEP). (2015). *Green Economy Sector Study on Agriculture in Kenya.* Nairobi-Kenya: UNEP. [Online]. Available at: http://drustage.unep.org/greeneconomy/sites/unep.org.greeneconomy/files/publications/kenya_agriculture_final.pdf [Accessed 3 October 2016].

United States Geological Survey. (2017). Management: Manage Quality. [Online]. Available at: http://www2.usgs.gov/datamanagement/qaqc.php [Accessed 1 August 2017].

University of California, Los Angeles. (n.d). Data Mining: What is Data Mining? [Online]. Available at: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm [Accessed 9 March 2017].

University of Cape Town (UCT). (2016). UCT Libraries joins Network of Data and Information Curation Communities. [Online]. Available at: http://www.eresearch.uct.ac.za/news/uct-libraries-joins-network-data-information-curation-communities [Accessed 4 April 2017].

The University of Edinburgh. (2016). Information Services. [Online]. Available at: http://www.ed.ac.uk/information-services/research-support/data-management/why-manage-data [Accessed 1 April 2016].

University of Oregon Libraries. (n.d). Research Data Management. [Online]. Available at: https://library.uoregon.edu/datamanagement [Accessed 22 May 2017].

University of Pretoria. (2007). Policy for the Preservation and Retention of Research Data. [Online]. Available at: http://webcache.googleusercontent.com/search?q= cache:igaTSFfrAOEJ:www.up.ac.za/media/shared/Legacy/sitefiles/file/40/policyforthepreservationandretentionofresearchdata.doc+andcd=1andhl=enandct=clnkandgl=za [Accessed 2 July 2017].

The University of Queensland. (2017). Research Data Management: Data Description. [Online]. Available at: http://guides.library.uq.edu.au/research-data-management/data-description [Accessed 9 June 2017].

The University of Sheffield, (n.d). Why Manage your Data? [Online]. Available at: https://www.sheffield.ac.uk/library/rdm/benefits [Accessed 28 June 2017].

The University of Sheffield. (2016). Benefits of Good RDM. [Online]. Available at: https://www.sheffield.ac.uk/library/rdm/benefits [Accessed 1 April 2016].

The University of Sheffield. (2017). Preservation of your Data. [Online]. Available at: https://www.sheffield.ac.uk/library/rdm/preserving [Accessed 28 July 2017].

University of South Africa Library. (2017). Research Data Management: Data Protection, Security, Storage and Backup. [Online]. Available at: http://libguides.unisa.ac.za/c.php?g=400358andp=2721783 [Accessed 16 June 2017].

University of Southern California. (2016). Organizing your Social Sciences Research Paper: Theoretical Framework. [Online]. Available at: http://libguides.usc.edu/writingguide/theoreticalframework [Accessed 27 June 2016].

University of Tasmania. (2016). Research Data Management and Discovery: Data Disposal. [Online]. Available at: https://utas.libguides.com/researchdatamanagement/datadisposal [Accessed 10 October 2017].

The University of Western Australia. (2016). Metadata Standards. [Online]. Available at: http://guides.is.uwa.edu.au/c.php?g=325196andp=2178564 [Accessed 3 August 2016].

Vahed, A. (2014). Data Intensive Research Initiative of South Africa (DIRISA): A reinterpreted vision. [Online]. Available at: http://www.eresearch.ac.za/sites/default/ files/image tool/images/140/DIRISA_busPlan_eResConf.pdf [Accessed 25 September 2016].

Van den Eynden, V. (2013). Data Life Cycle and Data Management Planning. [Online]. Available at: https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr 2013.pdf [Accessed 21 September 2016].

Van den Eynden, V. and Brett, B. M. (2010). Benefits of Data Management and Sharing. [Online]. Available at: http://www.hsrc.ac.za/uploads/pageContent/2109/Benefits%20of%20Data%20Management%20and%20Sharing.pdf [Accessed 2 November 2017].

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. and Horton, L. (2011). *Managing and Sharing Data best Practices for Researchers.* London, UK: Data Archives.

Van Deventer, M. V. and Pienaar, H. (2015). Research Data Management in a Developing Country: A Personal Journey. *International Journal of Digital Curation*, 10(2), 33-47.

Van Teijlingen, E. R. and Hundley, V. (2001). The Importance of Pilot Studies. [Online]. Available at: http://www.sru.soc.surrey.ac.uk/SRU35.html [Accessed 6 October 2017].

Van Tuyl, S and Michalek, G. (2015). Assessing Research Data Management Practices of Faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication*, 3(3), eP 1258 .

Van Wyk, J. and Van der Walt, I. (2014). Going Full Circle: Research Data Management @ University of Pretoria. Presentation at eReseach Africa 2014 Conference, held at University of Cape Town, Cape Town, South Africa 23-27 November 2014. [Online]. Available at: http://www.eresearch.ac.za/sites/default/files/image _tool/images/140/Going_Full-Circle_RDM_UP_VanWyk_VanderWalt_26Nov_2014.pdf [Accessed 28 March 2016].

Walton, G. (2010). Data Curation and the Academic Library. *New Review of Academic Librarianship*, 16(1), 1-3.

Wambani, H. N. (2011). Improving Scientific Research at Kenya Agricultural Research Institute (KARI)-Kakamega Research Centre: A Research Methods Approach. (Masters Dissertation). [Online]. Available at: http://ruforum.org/sites/default/files/Hildah%20Nyende%20Wambani%20thesis.pdf [Accessed 25 April 2016].

Warden, R. (2010). The Internet and Science Communication: Blurring the boundaries. *Ecancer medical science*, 4, 1-8.

Warren, J. (2016). Strengthening Research through Data Sharing. *The New England Journal of Medicine*, 375, 401-403.

Wassenaar, D. R. (2006). Ethical Issues in Social Science Research. In. T. Blanche, M.K. Durrheim and D. Painter (eds.). *Research in Practice: Applied Methods for the Social Sciences.* Cape Town, Soth Africa: University of Cape Town.

Weaver, K. and Olson, J.K. (2006). Understanding Paradigms Used for Nursing Research. *Journal of Advanced Nursing*, 53(4), 459-469.

Weber, N., Palmer, C. and Chao, T. (2012). Current Trends and Future Directions in Data Curation Research and Education. *In Journal of Web Librarianship*, 6(4).

Wegner, T. (2000). *Quantitative Methods for marketing Decisions*. Cape Town, Kenwyn: Juta and co.

Western Libraries. (2012). Digital Preservation. [Online]. Available at: http://www.lib.uwo.ca/scholarship/digitalpreservation.html [Accessed 20 October 2016].

Western Libraries. (2013). What is Scholarly Communication? [Online]. Available at: http://www.lib.owo.ca/scholarship/scholarlycommunication.html [Accessed 20 October 2016].

Whitehead, M. and Bourne-Tyson, D. (n.d). Multi-Stakeholder Engagement in Research Data Management. [Online]. Available at: https://webcache.googleusercontent.com/search?q=cache:UxfMleJLe8MJ:https://portagenetwork.ca/wp-content/uploads/2016/06/IATUL2016Multi_Stakeholder_Engagement_in_RDM.pdf+andcd=1andhl=enandct=clnk [Accessed 30 March 2017].

Whyte, A. (2014). A Pathway to Sustainable Research Data Services: From Scoping to Sustainability. In G. Pryor, S. Jones, and A. Whyte. (Eds.). *Delivering research data management services*. London: Facet. 59–88.

Whyte, A. and Allard, S. (2014). How to Discover Research Data Management Service Requirements. DCC How-to Guides. Edinburgh: Digital Curation Centre. [Online]. Available at: http://www.dcc.ac.uk/how- [Accessed 7 July 2016].

Whyte, A. and Tedds, J. (2011). Making the Case for Research Data Management, DCC Briefing Papers. Edinburgh: Digital Curation Centre. [Online]. Available at: http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm [Accessed 23 September 2016].

Whyte, A. and Wilson, A. (2010). How to Appraise and Select Research Data for Curation. Edinburgh: Digital Curation Centre. [Online]. Available at: http://www.dcc.ac.uk/resources/how-guides/appraise-select-data [Accessed 23 June 2016].

Wilson, N. and McLean, S. (1994). *Questionnaire Design: A Practical Introduction*. Newtown Abbey, Co. Antrim: University of Ulster Press.

Witt, M. (2008). Institutional Repositories and Research Data Curation in a Distributed Environment. *Library Trends*, 57(2), 191-201.

Woolfrey, L. (2014). University of Cape Town (UCT) Research Data management Policy Project: Report. [Online]. Available at: https://www.datafirst.uct.ac.za/images/docs/ 20140307-uct-rsearch-data-management-woolfrey.pdf [Accessed 27 September 2016].

The World Agroforestry Centre. (2012). Research Data Management Training Manual. [Online]. Available at: http://worldagroforestry.org/research-methods/sites/default/files/Research%20Data%20Management%20Manual%20-%20web%20temp.pdf [Accessed 29 March 2016].

Yakel, E. (2007). Archives and Manuscripts Digital Curation. *OCLC Systems and Services International Digital Library*. 23(4), 335-340.

Yakel, E., Ixchel, M., Faniel, A. K. and Yoon, A. (2013). Trust in Digital Repositories. *International Journal of Digital Curation*, 8 (1): 143-156.

Yin, R.K. (1989). *Case Study Research Design and Methods*. Newbury: Sage.

Yin, R.K. (2009). *Case Study Research: Design and Methods*. 4th (ed.). Thousand, Oaks: Sage.

Yiotis, K. (2008). Electronic Theses and Dissertation (ETD) Repositories: What are they? Where do they come from? How do they work? *OCLC Systems and Services: International Digital Library Perspectives*, 24, 101-115.

APPENDICES

APPENDIX 1: Interview schedule for directors of agricultural research institutes

Introduction:

I am a PhD student at the University of KwaZulu-Natal, South Africa at the School of Social Sciences, Information Studies Program. I am conducting research as part of the requirements for the award of a PhD degree in information studies. The topic of the research is "**Research Data Management practices in Kenya's Agricultural Research Institutes"**

The study will focus on Research Data Management (RDM) practices in Kenya's agricultural research institutes with the view to suggesting interventions to improve management, sharing and reuse of agricultural research data. The research will gather data on legal, policies and regulation affecting RDM; capture, appraisal, description, preservation, access, use and reuse of research data; the knowledge, skill and training requirements for RDM; level of ICT preparedness; and collaborative partnerships influencing RDM in agricultural research institute.

I kindly request your participation in this interview to enable me collect data that will address the research problem under investigation. Your responses will be treated with strict confidentiality, and will only be used for the stated academic purpose. Your contribution is highly appreciated. Please do not hesitate to contact me or my supervisor for clarification on any aspect of this interview.

Thank you in advance for your time and cooperation.

Emily Jeruto Ng'eno
PhD student (University of KwaZulu-Natal, South Africa)
E-mail: 216056748@stu.ukzn.ac.za
Mobile No.: +254 721484263
Supervisor: Prof. Stephen Mutula
Email: mutulas@ukzn.ac.za

A. Background information

1. Name of the research institute……………………………………………
2. Current designation ………………………………………………………….
3. Gender: Male […..] Female […..]
4. Age 26-30 […..] 31-35 […..] 36-40 […..] 41-45 [.....]
   46-49 […..] 50-above […..]
5. Field of specialization………………………………………………………
6. Please indicate how long you have worked in the current position

   Less than 1years […..] 1-5 years […..] 5-10 years […..]
   10-15 […..] Above 15 years [.....]

7. Highest academic qualification attained

   Diploma […..] Higher Diploma […..] Bachelors Degree […..]

   Masters Degree […..] PhD […..]

   Other(s) specify……………………………..…………………………………….

8. What is the mandate of the agricultural research institute you are affiliated to? ....
   ………………………………………………………………………………………

B. The availability or absence of legal, policy and regulations affecting the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes

1. What legal frameworks govern Research Data Management (RDM) practices in your institute?
2. Does your research institute have a formal RDM policy?
3. Which areas of RDM does the institutional policy cover e.g. data curation, open access, sharing, reuse, human capability, technical infrastructure and collaborations?
4. How is Intellectual Property (IP) managed with reference to RDM in your research institute?
5. What quality assurance and control measures are in place for RDM in your institute?

C. Research data capture, appraisal, description, preservation, accessibility and reuse in Kenya's agricultural research institute.

1. How is the research data captured, appraised, described, preserved, accessed and reused in your research institute?
2. What kinds of tools are available for research data capture, appraisal, description, preservation and access?
3. What importance does your research institute attach research data capture, appraisal, description, preservation, access, reuse and sharing?
4. Why do you reuse research data in your institute?
5. With whom does your research institute share research data generated and for what purpose?

D. RDM knowledge, skills and training requirements needed to capture, appraise, describe, preserve, access and reuse its research data

1. Who is responsible for research data capture, appraisal, description, preservation, and access in your research institute?
2. What are the competencies (knowledge and skill) required in research data capture, appraisal, description, preservation, access and reuse in your research institute?
3. What are the training needs for research data capture, appraisal, description, preservation, access and reuse in your research institute?
4. What strategies are in place to ensure research data are retained when researchers leave the research institute?
5. What support does your research institute provide researchers, librarians, IT specialists and RDM staff to enhance RDM?
6. What RDM literacy and advocacy programs are available in your research institute?
7. What are the contributions of library, and IT departments toward RDM?

E. Level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse of research data

1. What range of ICT(s) tools and equipment is available for data capture, appraisal, description, preservation and access in your research institute?
2. What research data services facilitated by ICT(s) are available in your research institute for research data capture, appraisal, description, preservation, access and reuse in your research institute?
3. What security measures are in place to protect research data during capture, appraisal, description, preservation and access in your research institute?
4. What is the level of ICT(s) infrastructure preparedness with regard to RDM in your institute?

F. Collaborative partnerships influencing the capture, appraisal, description, preservation, access and reuse of research data

1. Please explain any collaborative partnerships that exist within your research institute, across disciplines of research institute, with other research institute, at a national level or at the international level in the capture, appraisal, description, and preservation of research data?

3. What do you think is the impact of collaborative partnerships on:
   (a) Data open access? ………………………………………………………….
   (b) Data use and reuse? ………………………………………………………..
   (c) Data transformation?...
   (d) Data sharing? ……………………………………………………………….

4. What benefits accrue from your institution's collaborative partnerships in the capture, appraisal, description, and preservation of research data?

6. What is the source of your RDM budget?

9. (a) What are the benefits of RDM?
   (b) What challenges are faced during RDM in the research institute?
   (c) What suggestions can you make that can enhance RDM and address the challenges identified in 9b above?

Thank you for your time and cooperation

APPENDIX 2: Interview schedule for heads of research

Introduction:

I am a PhD student at the University of KwaZulu-Natal, South Africa at the School of Social Sciences, Information Studies Program. I am conducting research as part of the requirements for the award of a PhD degree in information studies. The topic of the research is "**Research Data Management practices in Kenya's Agricultural Research Institutes"**

The study focuses on Research Data Management (RDM) practices in Kenya's agricultural research institutes with the view to suggesting interventions to improve management, sharing and reuse of agricultural research data. The research will gather data on legal, policies and regulation affecting RDM; capture, appraisal, description, preservation, access, use and reuse of research data; the knowledge, skill and training requirements for RDM; level of ICT preparedness; and collaborative partnerships influencing RDM in agricultural research institute.

I kindly request your participation in this interview to enable me collect data that will address the research problem under investigation. Your responses will be treated with strict confidentiality, and will only be used for the stated academic purpose. Your contribution is highly appreciated. Please do not hesitate to contact me or my supervisor for clarification on any aspect of this interview.

Thank you in advance for your time and cooperation.

Emily Jeruto Ng'eno

PhD student (University of KwaZulu-Natal, South Africa)

E-mail: 216056748@stu.ukzn.ac.za

Mobile No.: +254 721484263

Supervisor: Prof. Stephen Mutula

Email: mutulas@ukzn.ac.za

A. Background information

1. Name of the research institute………………………………………………
2. Current designation ……………………………………………………………
3. Gender: Male […..] Female […..]
4. Age 26-30 […..] 31-35 […..] 36-40 […..] 41-45
   46-49 […..] 50-above […..]
5. Field of specialization……………………………………………………………
6. Please indicate how long you have worked in your current post in the research institute?
   Less than 1years […..] 1-5 years […..] 5-10 years […..]
   10-15 […..] Above 15 years [.....]
7. Highest academic qualification attained
   Diploma […..] Higher Diploma […..] Bachelors Degree […..]
   Masters Degree […..] PhD […..]
   Other(s) specify…………………………………………………………………
8. What is the role of the research unit? ....
9. What is the number of researchers under your jurisdiction in the research unit?
10. What duties do they perform? ………………………………………………………..

B. The availability or absence of legal, policy and regulations affecting the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes.

1. What legal frameworks are in place to govern Research Data Management (RDM) in your research unit?
2. What policies and regulations are in place to facilitate research data capture, appraisal, description, preservation, access, sharing, reuse, human capability, technical infrastructure and collaborations?
3. How is Intellectual Property (IP) managed with reference to RDM in your research institute?
4. What quality assurance and control measures are in place for RDM in your institute?

C. Research data capture, appraisal, description, preservation, access and reuse in Kenya's agricultural research institute.

1. How is the research data captured**,** appraised, described, preserved, accessed and reused?
2. What facilities are available to capture, appraise, describe, preserve, and access research data for reuse?
3. In what format is the research data captured appraised, described, preserved, and accessed and reused?
4. What are the benefits of using and re-using research data?
5. (a) What type of research data is usually shared within and outside the institute?

   (b) What mechanisms are provided for sharing research data?

   (c) With whom do you share research data?
6. What is the role of research unit in data capture, appraisal, description, preservation, access and reuse?

D. RDM knowledge, skills and training requirements for capturing, appraising, describing, preserving, accessing and re-using research data

1. Who is responsible for research data capture, appraisal, description, preservation, accessibility and reuse in your institute?
2. What type of knowledge and skills are required to handle RDM activities in your research institute?
3. What are the RDM training needs for your research institute?
4. What strategies are in place to ensure researchers leave behind research data when they disengage with the research institute?
5. What kind of support/services does your research department/institute provide to researchers to ensure effective data capture, appraisal, description, preservation, access and reuse
6. What are the roles of librarians and IT specialists in RDM?

E. Level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse research data

1. What types of ICT tools and equipment are used by the research institute in research data capture, appraisal, description, preservation, access and reuse?

2. What research data services (i.e. web portals, intranets) are available in your research institute for RDM?
3. What security measures are in place to protect research data from unauthorized access during data capture, appraisal, description, preservation, access, reuse and sharing?
4. What is the level of ICT(s) infrastructure preparedness for RDM?

F. Collaborative partnerships in the capture, appraisal, description, preservation, access and reuse of research data

1. What collaborative partnerships exist within and outside your research institute for RDM?
2. What has been the impact of collaborative partnerships in the research institute on:
   (a) Open access? ……………………………………………………………….
   (b) Use and reuse of data? …………………………………………………………..
   (c) Data transformation……………………………………………………………
   (d) Data sharing? ………………………………………………………………….
3. What are the benefits accrued from collaborative partnerships?
4. How is RDM funded in your institute and what conditions come with such funding?
5. (a) What are the benefits of RDM?
   (b) What challenges are faced during RDM in your research institute?
   (c) What suggestions can you make to address the challenges in 5b above?

Thank you for your time and cooperation

APPENDIX 3: Interview schedule for heads of IT and librarians

I am a PhD student at the University of KwaZulu-Natal, South Africa in the School of Social Sciences, Information Studies Program. I am conducting research as part of the requirements for the award of a PhD degree in information studies. The topic of the research is "**Research Data Management practices in Kenya's Agricultural Research Institutes"**

The study will focus on Research Data Management (RDM) practices in Kenya's agricultural research institutes with the view to suggesting interventions to improve management, sharing and reuse of agricultural research data. The research will gather data on legal, policies and regulation affecting RDM; capture, appraisal, description, preservation, access, use and reuse of research data; the knowledge, skill and training requirements for RDM; level of ICT preparedness; and collaborative partnerships influencing RDM in agricultural research institute.

I kindly request your participation in this interview to enable me collect data that will address the research problem under investigation. Your responses will be treated with strict confidentiality, and only be used for the stated academic purpose. Your contribution is highly appreciated. Please do not hesitate to contact me or my supervisor for clarification on any aspect of this interview.

Thank you in advance for your time and cooperation.

Emily Jeruto Ng'eno
PhD student (University of KwaZulu-Natal, South Africa)
E-mail: 216056748@stu.ukzn.ac.za
Mobile No.: +254 721484263

A. Background information

1. Name of the research institute………………………………………………
2. Current designation …………………………………………………………………
3. Gender: Male […..] Female […..]
4. Age 26-30 […..] 31-35 […..] 36-40 […..] 41-45 [.....]

   46-49 […..] 50-above […..]
5. Field of specialization…………………………………………………………………..
6. Please indicate how long you have worked at the research institute

   Less than 1years […..] 1-5 years […..] 5-10 years […..]
   10-15 […..] Above 15 years [.....]

7. Highest academic qualification attained

   Diploma […..] Higher Diploma […..] Bachelors Degree […..]
   Masters Degree […..] PhD […..]

   Other(s) specify……………………………………………………………………….

8. What is the role of your department/section in the research institute? ....

…………………………………………………………………………………………….

B. The availability or absence of Legal, policy and regulations affecting the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes.

1. What legal frameworks govern research data management (RDM) in your institute?
2. What policies and regulations does your department/section have for RDM and what do they cover with regard to data curation, open access, sharing, reuse, human capability, technical infrastructure and collaborative partnerships?
3. How is Intellectual Property (IP) rights regarding RDM?
4. What quality assurance and control measures are in place during for RDM in your department?

C. Research data capture, appraisal, description, preservation, accessibility and reuse in Kenya's Agricultural research institute.

1. How is research data captured, appraised, described, preserved, accessed and reused in your institute?
2. What types support are offered by your department/section to enhance research data capture, appraisal, description, preservation, access and reuse in your research institute?
3. What is the role of the department/section in research data capture, appraisal, description, preservation, access and reuse in your research institute?

D. RDM knowledge, skills and training requirements needed to capture, appraise, describe, preserve, and make accessible for reuse its research data

1. Who is responsible for research data capture, appraisal, description, preservation, access and reuse in your research institute?
2. What are the competencies (knowledge and skill) required in research data capture, appraisal, description, preservation, access and reuse in your research institute?
3. What methods are used in developing staff capability for research data capture, appraisal, description, preservation, and access and reuse in your research institute?
4. What data literacy and advocacy programme do you offer to facilitate research data capture, appraisal, description, preservation, access and reuse in your research institute?

E. Level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse of research data

1. What range of ICT(s) tools and equipment does the IT/Library department use in
    - (a) Research data capturing? ……………………………………………………….
    - (b) Research data appraisal? ……………………………………………………….
    - (c) Research data description? ……………………………………………………..
    - (d) Research data preservation? ……………………………………………………
    - (e) Research data accessing? ………………………………………………………
    - (f) Research data sharing? …………………………………………………………
    - (g) Others, (please specify) ……………………………………………………

2. What research data services does your department offer for research data capture, appraisal, description, preservation, access and reuse in your research institute?
3. What security measures are in place to protect research data from unauthorized access during data capture, appraisal, description, preservation, access and reuse?
4. What is the level of ICT(s) infrastructure preparedness with regard to RDM?

F. Collaborative partnerships influencing the capture, appraisal, description, preservation, access and reuse of research data

1. What type of collaborative partnerships on RDM does your department/section have within and outside the research institute?
2. What benefits accrue from collaborative partnerships with regard to RDM?
3. (a) What are the benefits of RDM?
   (b) What challenges are faced during RDM by the research institute?
   (c) How can the challenges faced in 4b above be addressed?

Thank you for your time and cooperation

APPENDIX 4 : Survey questionnaire for researchers

Dear Respondent

I am a PhD student at the University of KwaZulu-Natal, South Africa at the School of Social Sciences, Information Studies Programme. I am conducting research as part of the requirements for the award of a PhD degree in information studies. The topic of the research is "**Research Data Management practices in Kenya's Agricultural Research Institutes"**

The study focuses on research data management (RDM) practices in Kenya's agricultural research institutes with the view to suggesting interventions to improve management, sharing and reuse of agricultural research data. The research will gather data on legal, policies and regulation affecting RDM; capture, appraisal, description, preservation, access, use and reuse of research data; the knowledge, skill and training requirements for research data management; level of ICT preparedness; and collaborative partnerships influencing RDM in agricultural research institute.

I kindly request your participation in this survey to enable me collect data that will address the research problem under investigation. Your responses answers will be treated with strict confidentiality, and will only be used for the stated academic purpose. Your contribution is highly appreciated. Please do not hesitate to contact me for clarification on any aspect of this interview.

Thank you in advance for your time and cooperation.

Emily Jeruto Ng'eno
PhD student (University of KwaZulu-Natal, South Africa)
E-mail: 216056748@stu.ukzn.ac.za
Mobile No.: +254 721484263
Supervisor: Prof. Stephen Mutula
Email: mutulas@ukzn.ac.za

A. Background information

1. Name of the research institute………………...
2. Gender: Male […..] Female […..]
3. Age: 26-30 […..] 31-35 […..] 36-40 […..] 41-45 […..]
   46-49 […..] 50-above […..]
4. Field of specialization…………………………………………………………..………
5. Please indicate how long you have worked at the institute in years

   Less than 1years […..] 1-5 years […..] 5-10 years […..]
   10-15 […..] Above 15 years [.....]

9. Highest academic qualification attained
   Diploma […..]   Higher Diploma […..]   Bachelors Degree […..]

   Masters Degree […..]   PhD […..]

   Other(s) (please specify)…………………………………………………………

B. The availability or absence of legal, policy and regulations affecting the capture, appraisal, description, preservation, access and reuse of research data by Kenya's agricultural research institutes

1. Are there legal frameworks that govern research data management (RDM)?

   […..] Yes   […..] No

   If yes, explain……………………………………..

   If no, explain………………………………………………………………….

2. What policies and regulations are available to facilitate the capture, appraise, description, preservation, access, reuse and sharing of research data by Kenya's agricultural research institutes? (*Please rate your answer on a five point scale by indicating your level of agreement by ticking the appropriate response)*
   Strongly Agree (SA), Agree (A), Undecided (U), Disagree(D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
The Research Institute has Research Data Management (RDM) Policy					
The Research Institute has Research Data Management Policy governing:					
• Research data capture					
• Research data appraisal					
• Research data description					
• Research data preservation					
• Research data access					
• Research data use and reuse					
• Research data sharing					
• Knowledge, skills and training					
• Technical infrastructure					
• Collaborative partnerships					
The Research Institute owns the right to research data					
The Researcher owns the right to research data created					
Intellectual property (copyright, patents, trademarks) affects RDM					
There is quality assurance and control measures in place during:					
Research data capture					
Research data appraisal					

Research data description					
Research data preservation					
Research data access					
Research data reuse					

C. Research data capture, appraisal, description, preservation, accessibility and reuse in Kenya's agricultural research institute

Capturing research data

1. Please explain the method of capturing research data?..

....

2. Please rate your answer on capturing research data on a five point scale by indicating your level of agreement by ticking the appropriate response.
   Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
In what formats do you generate your research data					
Audio					
Images					
Spreadsheets					
Video					
Data-Statistical (SAS, SPSS)					
Database					
Scanned documents					
Web					
Computer Aided Design (CAD)					
Geographic Information Systems (GIS)					
Data XML					
Others (please specify)					
Approximate amount of research data your institute generates in every research project (tick where appropriate)					
1-500 gigabytes (GB)					
500-1000 gigabytes GB					
1-500 terabytes (TB)					
500-1000 TB					
1-500 Petabytes					
>500 PB					
Don't know					

Appraisal of research data

Please rate your answer on appraisal of research data on a five point scale by indicating your level of agreement by ticking the appropriate response.
Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
Research data appraisal checklist in your institute					
Uniqueness					
Repeatability					
Scientific/historical value					
Complementary/added value					
Reuse value					
Substantiveness					
Access					
Volume					
Cost-effectiveness					
Tools used to guide the appraisal of research data					
Appraisal and selection policy					
Research data management policy					
Research institute's policy					
Others (please specify)					

Description of research data (i.e. metadata)

Please rate your answer on description of research data (i.e metadata) on a five point scale by indicating your level of agreement by ticking the appropriate response.

Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
Describing each of your research datasets					
The only description will be the filenames on my hard drive.					
I will use handwritten notes in my lab notebook after the experiments have been completed.					
I will describe the data using the column and row labels in my spreadsheets after the data have been analysed.					
I will create descriptive metadata for each dataset and will save these descriptions with my datasets on my hard drive.					

Creating or capturing descriptive metadata

Instrument metadata are automatically included in each data file [ ]

I will create a title and short textual description for each dataset when submitting the dataset to my research institute data repository [ ]

My data descriptions will be saved in spreadsheets or word processor documents [ ]

I will create rich metadata by recording data at the time of capturing using a metadata entry form to ensure I don't miss any essential information. This metadata file will be saved locally with my dataset, and eventually will be deposited with the dataset when it is submitted to a data repository [ ]

Storage, Backups and Preservation of Research Data

Please rate your answer on storage, backups and preservation of research data (i.e metadata) on a five point scale by indicating your level of agreement by ticking the appropriate response.

Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
Storage of your research data in short term, after acquisition					
On my laptop					
On the computer connected to the server					
On my research group's Data storage file					
CDs or DVDs					

Storage media of your research data in long term

Hard drive of the instrument which generates the data	[ ]
PC hard drive	[ ]
External hard drive	[ ]
Departmental server	[ ]
CD/DVD	[ ]
USB flash drives	[ ]
Repository	[ ]
Internet-based storage (e.g., cloud or grid storage)	[ ]
Others (please specify)	

Who is responsible for day-to-day management, storage and backup of the data arising from your research?	**SA**	**A**	**U**	**D**	**SD**
Myself					
My research group's data manager.					
Department					
IT staff who manage our research group's DataStage (cloud storage)					
Librarian					
The research unit					
Others (please specify)					
Frequency of making backups for the research data					
Hourly					
Daily					
Weekly					
Monthly					
Annually					
Never					
Others (please specify)					

For how long is research data kept in your research institute

<1 year	[ ]
1–5 years	[ ]
5–10 years	[ ]
>10 years	[ ]
Don't know	[ ]

Please explain why you plan on keeping the data for this amount of time					
Who decides which of your research data are worth preserving?	**SA**	**A**	**U**	**D**	**SD**
Myself alone					
Myself, in consultation with my research supervisor					
My research supervisor alone					
Institutional policy					
Funders requirement					
When is your research data moved to a secure archive for long-term preservation					
Upon completion of each set of experiments					
When my research group leader decides it is appropriate					
Immediately after publication of my paper					
Upon submission of the paper, so that the data are available for reviewers					
Determined by RDM policy					
Where is your research data archived for long-term preservation					
Selected data will be included in the figures and tables of research papers published by my research group, but we have no plans to archive and publish the full datasets					
As supplementary files attached to my journal articles on the publisher's web site.					
In the research institute's data Archive server					
In the research institutes Data Bank					
Transfer of your research datasets (by physical or electronic method) for long-term archiving, under the curatorial care e.g. a data repository					
Using physical hard drives					
By e-mailing files to our IT department or Librarian					
By transferring to Web-based selected data repository					
By use of a local data management system such as DataStage that can automatically package and submit data files to the selected repository					
Who will be responsible for your research data, once you have left your present research institute?					
I'll take my data with me and manage it					
My supervisor will be responsible					
The research institute will assume responsibility for the data I have chosen to preserve in its data archive					

Research Data access, sharing and re-using

Please rate your answer on data access, reuse and sharing of research data on a five point scale by indicating your level of agreement by ticking the appropriate response.

Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).

Statement	SA	A	U	D	SD
Time taken to halt your research data before it is published					
We allow immediate public access to the data					
For one year, to permit us to exploit our hard-won research results					
Until our journal article has been published					
Others (please specify)					
Restricting public access to your research data					
We intend to make a patent application thus must avoid prior disclosure					
The research data are confidential because of the arrangement my research group has made with the commercial partner sponsoring our research					
Confidential, proprietary or classified research data					
Intellectual property concerns					
Lack of appropriate tools for accessing or publishing data					
Other (please specify)					

Who currently accesses your research data?

Group researchers [ ]
Other researcher in the research institute [ ]
Researchers from other research institutes [ ]
Open to public [ ]

Availing research data to others through open access

Through institutional repositories **[ ]**
Through published literature **[ ]**
Through conferences, seminar and workshops **[ ]**
Through the media **[ ]**

With whom are you sharing your research data before publication of any papers arising from their interpretation	SA	A	U	D	SD
My research supervisor					
Members of my research group and trusted external collaborators					
Research sponsors					
General public					
Everyone, by publishing the data online					
Others (please specify)					
Methods used to share your all or part of research data					
Collaborative web space (wiki, blog, Google Docs)					
Data portal or database driven web site					
Deposit them with a specialized data centre					

Depositing them in an institutional repository					
Submitting them to a journal to support publication					
E-mail					
External storage device (USB drive, CD/DVD)					
Hard copy					
Don’t share data					
Other (please specify)					
Benefits of sharing research data					
Promote innovations and potential new data uses					
Encourages scientific enquiry and debate					
Reduce the cost of duplicating data collection					
Enables scrutiny of research findings					
Increases the impact and visibility of research					
Leads to new collaborations between data users and data creators					

Challenges encountered during research data sharing

Legal issues relating to patient privacy **[ ]**
Need for incentives **[ ]**
Fears regarding misuse of shared data **[ ]**
Building trust **[ ]**
Others (please specify) **[ ]**

Reasons for using and re-using research data	**SA**	**A**	**U**	**D**	**SD**
Avoid duplication					
Reduce the cost of doing research					
Re-analysis of data can lead to powerful insights					
Encourages scientific enquiry					
Promote innovations					
Potential new data uses					
Way of making your research data available for use and reuse					
Publication					
Citation					
Sufficient metadata describing how the data has been specified, collected, analysed and transformed					
Research data licensing approaches					
Who is responsible for Research Data Management in your research institute?					
IT staff within research institute					
Librarian					
Collaborative responsibility or research group					
External research partners					
Third-party data center					
Other (please specify)					

D. RDM knowledge, skills and training requirements needed to capture, appraise, describe, preserve, access and reuse its research data

1. Please rate your answer on RDM knowledge, skills and training requirements on a five point scale by indicating your level of agreement by ticking the appropriate response. **Highly Skilled (HS), Skilled (S), Neutral (N), Less Skilled (LS), None Skilled (NS).**

Statement	HS	S	N	LS	NS
Type of knowledge you posses for research data capture, appraisal, description, preservation, access and reuse in your research institute					
Data Curation (capture, appraisal, description)					
Preservation					
Open access					
Sharing					
Others, (please specify)					
Type skills you posses for research data capture, appraisal, description, preservation, access and reuse in your research institute					
Metadata skills					
Searching and retrieval skills					
Data capturing skills					
Preservation skills					
Sharing skill					
Tools and Technologies (cloud computing, statistical analysis)					
Collaboration and communication (engaging with other researchers, the public)					
Others, (please specify)					
Areas in Research Data Management you feel there is a skills gaps					
Developing metadata schema					
Use of Information and Communication Technologies (ICTs) tools and equipments					
Data curation (data capture, appraisal, description, preservation, access and reuse)					
Complying with the various mandates of funders					
Use of institutional repository					
Others, (please specify)					
Rate the type of training areas that are highly needed in your department on a five point scale by indicating your level of agreement by ticking the appropriate response. **Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD).**					
	SA	**A**	**U**	**D**	**SD**
Metadata training					

Data curation (data capture, appraisal, description, preservation, access and reuse)					
Information and communication technologies (ICTs)					
Legal and Ethical consideration in RDM					
Funder requirement for research data					
Data management plans					
Data sharing					
Security and storage of data					
Collaborative partnerships					
Others (please specify)					

2. How would you rate the quality of the following research data services offered by the library and IT department? (*Please rate your answer on a five point by indicating your level of agreement by ticking the appropriate response)*
Very Satisfied (VS), Satisfied (S), Neutral (N), Dissatisfied (D), Very Dissatisfied (VD).

Statement	VS	S	N	D	VD
Directly participating with Librarians and IT specialist on RDM					
Providing reference and advisory support to researchers on RDM					
Creating web guides and finding aids for data repositories					
Providing technical support for research data service systems (repository, web portals, creating metadata, RDM system, access and discovery systems)					
Provision of RDM literacy programs					
Coordinating RDM advocacy programs between the researchers and the stakeholders					

3. Have you undertaken any data literacy programs in the research institute

[…..] Yes […..] No

If yes, what data literacy programs have you under taken…………………………..

If no, give reason……………………………………………………………………..

E. Level of ICT preparedness for the capture, appraisal, description, preservation, access and reuse of research data

- What ICT tools and equipment do you use in:
  - Research data capture…………………………………………………………
  - Research data appraise……………………………………………………..............
  - Research data describe………………………………………………………..
  - Research data preserve………………………………………………………..

- Research data access…………………………………………………………….
- Research data sharing……………………………………………………………..

- (a) What types of Research Data Management (RDM) software(s) do you use in research data capture, appraisal, description, preservation, access, reuse and sharing in your research institute?...

  ………………………………………………………………………………………………
  ……………………………………………………………………………………………

  (b) What are the strengths of mentioned software(s) in 2a above……..........................

  ....
  ………………………………………………………………………………………

  (c)What are the weaknesses of the same software(s)…………………………………

  ……………………………………………………………………………………...

- What is the frequently used data security?

  […..] Physical security

  […..] Network security

  […..] Encryption

  Others, (please specify)……..

F. Collaborative partnerships influencing the capture, appraisal, description, preservation, access and reuse of research data

4. How has research data capture, appraise, description, preservation, access and reuse in Kenya's agricultural research institute influenced collaborative partnerships? Rate your answer on a five point scale ranging from:

Very Satisfied (VS), Satisfied (S), Neutral (N), Dissatisfied (D), Very Dissatisfied (VD).

Statement	VS	S	U	D	VD
Collaboration within the discipline/sector					
How satisfied are you with collaboration with departmental research groups					
How satisfied are you with collaboration across research group between organizations					
How satisfied are you with collaborations organized at national level					
How satisfied are you with international collaborations and consortia					
Collaboration and interaction across discipline					
How satisfied are you with individual researchers occasionally collaborating outside their discipline					
How satisfied are you with interactions across disciplines collaborating through joint conferences or publications					

How satisfied are you with formal collaboration between research groups from different disciplines					
Collaboration with the public					
How satisfied are you with informational or participative media programs organized to engage the public					
Partnership: Geographical scale of funding for research					
How satisfied are you with research funded internally or through grants from regional agencies					
How satisfied are you with research funded by government					
How satisfied are you with research funding by international bodies					
Public-private partnerships					
How satisfied are you with informal partnerships with public and private industry but no funding involved					
How satisfied are you with research co-funded by public-private industry					
How satisfied are you with established formal co-investment partnerships running long-term multi-phase research					

5. Do you have a Data management plan?

[…..] Yes […..] No

- If yes, what does it entail? ....
- If no, please give reason ....

6.What are the benefits of RDM ....

....

....

7.What are the challenges of RDM in your research institutes?............................

....

....

8.What suggestions would you recommend to mitigate the challenges mentioned above?...

....

Thank you for your time and cooperation

APPENDIX 5: Ethical Clearance

UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

12 December 2016

Ms Emily Jeruto Ng'eno 216056748
School of Social Sciences
Pietermzritzburg Campus

Dear Ms Ng'eno

Protocol reference number: HSS/2083/016D
Project title: Research Data Management Practices in Kenya's Agricultural Research Institutes

Full Approval – Expedited Application

In response to your application received 28 November 2016, the Humanities & Social Sciences Research Ethics Committee has considered the abovementioned application and the protocol has been granted **FULL APPROVAL.**

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment /modification prior to its implementation. In case you have further queries, please quote the above reference number.

PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.

The ethical clearance certificate is only valid for a period of 3 years from the date of issue. Thereafter Recertification must be applied for on an annual basis.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

....

Dr Shenuka Singh (Chair)
Humanitities & Social Scinces Research Ethics Committee

/pm

Cc. Supervisor: Prof S Mutula
Cc. Academic Leader: Professor Maheshvari Naidu
Cc. School Administrator: Ms Nancy Mudau & Lukong Stella Shulika

Humanities & Social Sciences Research Ethics Committee
Dr Shenuka Singh (Chair)
Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Telephone: +27 (0) 31 260 3587/8350/4557 **Facsimile**: +27 (0) 31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za / mohunp@ukzn.ac.za
Website: www.ukzn.ac.za

1910 - 2010
100 YEARS OF ACADEMIC EXCELLENCE

APPENDIX 6: Informed Consent Letter

University of KwaZulu-Natal Library

UNIVERSITY OF KWAZULU-NATAL

Medical Library
Private Bag X7
Congella
4013
Telephone: 031 -260-4373
Fax: 031- 260- 4426
Email: ngcobon15@ukzn.ac.za

9 June 2016

Dear Respondent

Informed Consent Letter

Researcher: Emily Jeruto Ng'eno
Institution; University of KwaZulu-Natal
Telephone number: +27 628161019
Email address: 216056748@stu.ukzn.ac.za

Supervisor: Prof. S. Mutula
Institution: University of KwaZulu-Natal
Telephone number: +27 712750109
Email address: mululas@ukzn.ac.za

I, Emily Jeruto Ng'eno, of University of KwaZulu-Natal, kindly invite you to participate in the research project entitled 'Research Data Management practices in Kenya's Agricultural Research Institutes'

This research project is undertaken as part of the requirements of the PhD, which is undertaken through the University of KwaZulu-Natal, Information Studies Department.

The purpose of this study is to examine Research Data Management (RDM) practices in the Kenya's agricultural research institutes. The study wants in general to gain a deep understanding of how research data generated through agricultural research institutes in Kenya is captured, organized, preserved, accessed, disseminated and reused.

Participation in this research project is voluntary. You may refuse to participate or withdraw from the research project at any stage and for any reason without any form of disadvantage. There will be no monetary gain from participating in this research project. Confidentiality and anonymity of records identifying you as a participant will be maintained by the Department of Information Studies, at the University of KwaZulu-Natal.

If you have any questions or concerns about participating in this study, please feel free to contact myself or my supervisor at the numbers indicated above.

It should take you about 15 minutes to complete the questionnaire.

Thank you for participating in this research project.

9th June 2016

Signature Date

I .... hereby consent to participate in the above study.

Name: .... Date: .... Signature: ....

Supervisor's details
Prof. Stephen Mutula
Institution: University of KwaZulu-Natal
Email address: mutulas@ukzn.ac.za

Student's details
Emily Jeruto Ng'eno
Institution: University of KwaZulu-Natal
Tel. No.: +254 721 484 263
Email address: ngenojeruto@gmail.com

APPENDIX 7: Request to NACOSTI to undertake research

Information Studies
School of Social Sciences
University of KwaZulu-Natal
Private Bag X01
Scottsville 3209, South Africa
Tel: +27 (0) 33 2605571
Fax: +27 (0) 33 2605092
mutulas@ukzn.ac.za

9th June, 2016

The Director,
National Commission for Science, Technology and Innovation (NACOSTI),
8th-9th Floor, Utalii House Off Uhuru Highway, Nairobi Kenya,
P.O. Box 30623, 00100,
Nairobi,
Kenya.

Dear Sir/Madam,

RE: REQUEST FOR PERMISSION TO UNDERSTAKE RESEARCH

Reference is made to the above subject.

I am a Kenyan duly registered PhD student in the Information Studies Programme at the University of KwaZulu-Natal, in South Africa. As part of the requirement for the award of the doctoral degree, I am undertaking research on a topic titled "**Research Data Management Practices in Kenya's Agricultural Research Institutes**". The research will target Kenya's agricultural research institutes namely: International Livestock Research Institute (ILRI); World Agroforestry Centre (ICRAF); Biotechnology Research Institute (BRI); Coffee Research Institute (CRI); Food Crops Research Institute (FCRI) and Genetic Resources Research Institute (GeRRI). Data collection instrument that will apply include interviews to directors of institutes, heads of research, heads of IT, librarians and archivists and questionnaires to researchers. This study will assist in proposing interventions to improve management, sharing and reuse of agricultural research output.

The purpose of this letter is to kindly request for permission from your office to carry out research at the said Kenya's agricultural research institutes within October, 2016 to May, 2017. Should need further clarification please contact me and/or my supervisor Prof. Stephen Mutula at mutulas@ ukzn.ac.za.

I shall be grateful for your assistance and cooperation.

Yours sincerely,

Emily Jeruto Ng'eno
Tel. +254 721484263
E-mail: ngenojeruto@gmail.com

APPENDIX 8: Authority from NACOSTI to carry out research

NACOSTI

NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY AND INNOVATION

Telephone:+254-20-2213471, 2241349,3310571,2219420
Fax:+254-20-318245,318249
Email:dg@nacosti.go.ke
Website: www.nacosti.go.ke
when replying please quote

9th Floor, Utalii House
Uhuru Highway
P.O. Box 30623-00100
NAIROBI-KENYA

Ref. No
NACOSTI/P/16/29174/13536

Date:
2nd September, 2016

Emily Jeruto Ngeno
University of Kwa-Zulu Natal
SOUTH AFRICA.

RE: RESEARCH AUTHORIZATION

Following your application for authority to carry out research on "***Research Data Management Practices in Kenya`s Agricultural Research Institutes,"*** I am pleased to inform you that you have been authorized to undertake research in **Kiambu, Nairobi and Trans Nzoia Counties** for the period ending **2nd September, 2017.**

You are advised to report to **the Director General, selected Agricultural Institutes, the County Commissioners and the County Directors of Education, Kiambu, Nairobi and Trans Nzoia Counties** before embarking on the research project.

On completion of the research, you are expected to submit **two hard copies and one soft copy in pdf** of the research report/thesis to our office.

BONIFACE WANYAMA
FOR: DIRECTOR-GENERAL/CEO

Copy to:

The Director General
Selected Agricultural Institutes.

The County Commissioner
Kiambu County.

The County Director of Education
Kiambu County.

THIS IS TO CERTIFY THAT:
MS. EMILY JERUTO NGENO
of UNIVERSITY OF KWA-ZULU NATAL,
0-30100 Eldoret,hasbeen permitted to
conduct research in *Kiambu, Nairobi,*
Transnzoia Counties

on the topic: *RESEARCH DATA MANAGEMENT PRACTICES IN KENYA`S AGRICULTURAL RESEARCH INSTITUTES*

for the period ending:
2nd September,2017

Permit No : NACOSTI/P/16/29174/13536
Date Of Issue : 2nd September,2016
Fee Recieved :Ksh 2000

....
Applicant's Signature

....
Director General
National Commission for Science, Technology & Innovation

CONDITIONS

1. You must report to the County Commissioner and the County Education Officer of the area before embarking on your research. Failure to do that may lead to the cancellation of your permit.
2. Government Officer will not be interviewed without prior appointment.
3. No questionnaire will be used unless it has been approved.
4. Excavation, filming and collection of biological specimens are subject to further permission from the relevant Government Ministries.
5. You are required to submit at least two(2) hard copies and one (1) soft copy of your final report.
6. The Government of Kenya reserves the right to modify the conditions of this permit including its cancellation without notice

REPUBLIC OF KENYA

NACOSTI

National Commission for Science, Technology and Innovation

RESEACH CLEARANCE PERMIT

Serial No.A 10814

CONDITIONS: see back page

APPENDIX 9: Request to undertake research in KALRO-(1st letter)

Information Studies
School of Social Sciences
University of KwaZulu-Natal
Private Bag X01
Scottsville 3209, South Africa
Tel: +27 (0) 33 2605571
Fax: +27 (0) 33 2605092
mutulas@ukzn.ac.za

UNIVERSITY OF KWAZULU-NATAL

9th June, 2016

The Director General,
Kenya Agricultural and Livestock Research Organization (KALRO),
Location: Kaptagat Rd, Loresho Nairobi Kenya,
P.O. Box 57811,
City Square, Nairobi, 00200,
Kenya.
Email: directorgeneral@kalro.org

Dear Sir/Madam,

RE: APPLICATION FOR RESEARCH DATA COLLECTION

Reference is made to the above subject.

Ms. Emily Jeruto Ng'eno is a duly registered PhD student in the Information Studies Programme at the University of KwaZulu-Natal, in South Africa. As part of the requirement for the award of the doctoral degree, she is undertaking a study on "Research Data Management Practices in Kenya's Agricultural Research Institutes". The study will target four (4) agricultural research institutes under KALRO namely Biotechnology Research Institute (BRI); Coffee Research Institute (CRI); Food Crops Research Institute (FCRI) and Genetic Resources Research Institute (GeRRI).

The purpose of this letter is to kindly request a written permission from your office to enable her collect data from the said research institutes. Possible dates for data collection are flexible within October, 2016 to May, 2017. The data will be collected through survey questionnaire and interviews. Your authorization to this request will be highly appreciated.

Yours sincerely,

Professor Stephen Mutula
Dean and Head: School of Social Sciences

APPENDIX 10: Authority letter from KALRO

KALRO
Kenya Agricultural & Livestock Research Organization

KENYA AGRICULTURAL & LIVESTOCK RESEARCH ORGANIZATION

When replying please quote:

Our Ref: KALRO/3/041

Date: 2nd August, 2016

Professor Stephen Mutula,
Dean & Head: School of Social Sciences,
University of Kwazulu-Natal,
King Edward Avenue,
Scottsville Pietermaritzburg,
Private Bag X01, Scottsville 3209,
SOUTH AFRICA.

Dear Prof. Mutula,

RE: RESEARCH DATA COLLECTION – EMILY JERUTO NGENO

Reference is made to your letter dated 18th July, 2016 concerning the above subject.

In response to your request for your student Ms. Emily Jeruto undertaking PhD in the Information Studies Programme at your University on the topic **"Research Data Management Practices in Kenya's Agricultural Research Institutes",** we appreciate the fact that you have chosen KALRO to undertake your research.

We do support research undertakings for academics which correlates to the ability to conduct efficient and innovative research. Therefore, in the process of the student undertaking her research she will adhere to the Institutes' research regulations and should be able to share the research findings with the organization. In respect to this, the student is allowed to collect data from the 4 KALRO Institutes namely Beef Research Institute, Coffee Research Institute, Food Crops Research Institute and Genetic Resources Research Institute.

Yours sincerely,

Eliud K. Kireger (PhD)
DIRECTOR GENERAL

cc: Institute Director, BRI
Institute Director, CRI
Institute Director, FRI
Institute Directo, GeRRI

KALRO HEADQUARTERS,
P.O. Box 57811-00200, Nairobi, KENYA. Tel: 254-020 4183301-20 / 254-020 4183720 Fax: 254-020 4183344
Website: www.kalro.org

APPENDIX 11: Request to undertake research-KALRO

Information Studies
School of Social Sciences
University of KwaZulu-Natal
Private Bag X01
Scottsville 3209, South Africa
Tel: +27 (0) 33 2605571
Fax: +27 (0) 33 2605092
mutulas@ukzn.ac.za

UNIVERSITY OF KWAZULU-NATAL

25th October, 2016

The Director General,
Kenya Agricultural and Livestock Research Organization (KALRO),
Location: Kaptagat Rd, Loresho Nairobi Kenya,
P.O. Box 57811,
City Square, Nairobi, 00200,
Kenya.
Email: directorgeneral@kalro.org

Dear Sir/Madam,

RE: APPLICATION FOR RESEARCH DATA COLLECTION

Reference is made to the above subject.

Ms. Emily Jeruto Ng'eno is a duly registered PhD student in the Information Studies Programme at the University of KwaZulu-Natal, in South Africa. As part of the requirement for the award of the doctoral degree, she is undertaking a study on "Research Data Management Practices in Kenya's Agricultural Research Institutes". The study had originally targeted four (4) agricultural research institutes under KALRO namely Biotechnology Research Institute (BRI); Coffee Research Institute (CRI); Food Crops Research Institute (FCRI) and Genetic Resources Research Institute (GeRRI) of which you kindly granted permission.

However, it has become necessary to expand in order to have a more elaborate and exhaustive research. We have chosen the additional two institutes namely, Dairy Research Institute and Tea Research Institute.

The purpose of this letter is to kindly request for another written permission from your office approving Emily's research for the additional two institutes for my attention and subsequent clearance to enable her collect the data. Possible dates for data collection are flexible within October, 2016 to May, 2017. The data will be collected through survey questionnaire and interviews. Your authorization to this request will be highly appreciated.

Yours sincerely,

Professor Stephen Mutula
Dean and Head: School of Social Sciences

APPENDIX 12: Additional authority letter from KALRO

Kenya Agricultural & Livestock Research Organization

KENYA AGRICULTURAL & LIVESTOCK RESEARCH ORGANIZATION

When replying please quote:

Our Ref: KALRO/3/041

Date: 9th November, 2016

Professor Stephen Mutula,
Dean & Head: School of Social Sciences,
University of Kwazulu-Natal,
King Edward Avenue,
Scottsville Pietermaritzburg,
Private Bag X01, Scottsville 3209,
SOUTH AFRICA.

Dear Prof. Mutula,

RE: RESEARCH DATA COLLECTION – EMILY JERUTO NGENO

Further to our letter dated 2nd August, 2016 concerning the above subject (copy attached for ease of reference) we note that you have selected two more KALRO Institutes namely Tea Research Institute (TRI) and Dairy Research Institute (DRI) in addition to the four earlier Institutes.

This is to confirm that we have no objection for your student Ms. Emily Jeruto undertaking PhD in the Information Studies Programme at your University on the topic **"Research Data Management Practices in Kenya's Agricultural Research Institutes",** to collect data from our two institutes as per your request. All the other terms of our previous letter remain the same.

By copy of this letter the Institute Directors are hereby requested to assist the student.

Yours sincerely,

Eliud K. Kireger (PhD)
DIRECTOR GENERAL

cc: Institute Director, TRI
Institute Director, DRI

KALRO HEADQUARTERS,
P.O. Box 57811-00200, Nairobi, KENYA. Tel: 254-020 4183301-20/ 254-020 4183720 Fax: 254-020 4183344
Website: www.kalro.org