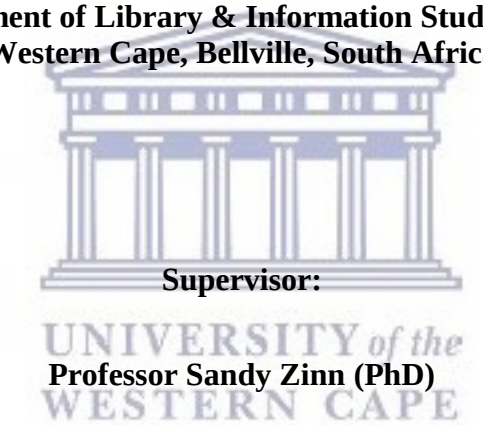
Research data management in public universities in Malawi

Winner Dominic Chawinga

BLIS, MLIS

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Library & Information Studies at the University of the Western Cape, Bellville, South Africa



Supervisor:

Professor Sandy Zinn (PhD)

17 May 2019

DECLARATION

I, **Winner Dominic Chawinga**, declare that:

(i) The research reported in this thesis, except where otherwise indicated, is my original work.

(ii) This thesis has not been submitted for any degree or examination at any other university.

(iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This thesis does not contain other persons' writings, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then: (a) their words have been re-written, but the general information attributed to them has been referenced; (b) where their exact words have been used, their writing has been placed inside quotation marks and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the sources being detailed in the dissertation and in the references sections.

Signed:  Date: 17 May 2019

Supervisor: Professor Sandy Zinn (PhD)

Signed:  Date: 17 May 2019

ABSTRACT

The emergence and subsequent uptake of Information and Communication Technologies has transformed the research processes in universities and research institutions across the globe. One indelible impact of Information and Communication Technologies on the research process is the increased generation of research data in digital format.

This study investigated how research data has been generated, organised, shared, stored, preserved, accessed and re-used in Malawian public universities with a view to proposing a framework for research data management in universities in Malawi. The objectives of the study were: to determine research data creation, sharing and re-use practices in public universities in Malawi; to investigate research data preservation practices in public universities in Malawi; to investigate the competencies that librarians and researchers need to effectively manage research data; and to find out the challenges that affect the management of research data in public universities in Malawi.

Apart from being guided by the Community Capability Model Framework (Lyon, Ball, Duke & Day, 2011) and Data Curation Centre Lifecycle Model (Higgins, 2008), the study was inspired by the pragmatic school of thought which is the basis for a mixed methods research enabling the collection of quantitative and qualitative data from two purposively selected universities. A census was used to identify researchers and librarians while purposive sampling was used to identify directors of research. Questionnaires were used to collect mostly quantitative and some qualitative data from 36 librarians and 187 researchers while interviews were conducted with directors of research. The Statistical Package for the Social Sciences was used to analyse the quantitative data by producing percentages, means, independent samples t-test and one-way analysis of variance. Thematic analysis was used to analyse the qualitative data.

Findings established that universities were involved in research activities which led to the creation of large quantities of research data. The study established that researchers rarely shared their data because of poor or the unavailability of data sharing infrastructure and absence of rewards or incentives. Although universities were aware of the benefits of data re-use, inaccessibility to re-usable data was a challenge. Generally, universities followed poor data preservation standards by using less dependable free standing devices such as external hard drives and personal laptops to store and back up data - these were highly susceptible to accidental damage and losses. Librarians and researchers had gaps in various key areas of

research data management due to failure by universities to conduct training workshops. The study exposed various factors that thwarted research data management activities including lack of data infrastructure, lack of skills, lack of incentives and recognition, lack of collaboration in research data management activities, and absence of research data management policies. In an effort to foster successful research data management, the study advanced several recommendations that if taken into consideration by various research stakeholders, could propel research data management in Malawian universities.

In addition, based on the findings of the study, an integrated framework for examining and understanding research data management in Malawian universities was proposed and documented. The model is composed of five elements: collaboration, RDM policies, RDM rewards, infrastructure, and RDM competence.

Keywords: Digital Curation Centre Curation Lifecycle Model, Community Capability Model Framework**,** Librarians, Malawi, Health sciences, Research data management, Researchers.

ACKNOWLEDGEMENTS

My jittery journey in pursuit of a doctorate degree has finally seen light at the end of the slippery tunnel. Experiences I have accrued from this journey I cherish and I will ever. I would be remiss not to thank everyone who pulled my hands along the way.

First, Prof Zinn. An awesome supervisor, meticulous advisor and exceptional mentor who first engaged me in my Master's thesis and later encouraged me to proceed with doctoral studies. For the past three years, she has been patient with me and very supportive in every possible way to ensure I succeeded in my studies. Precisely, Prof Zinn has done a superlative work in mentoring me academically beyond imagination though no surprise to me.

On the Faculty of Arts (University of the Western Cape) side, the commonly unsung faculty representatives have been supportive and friendly. In particular, I would like to single out Mr Leslie Richards who selflessly helped me with nerve-wracking registration issues year in, year out! We have not met nor spoken voice-to-voice, but he has my great admiration and gratitude.

I further thank my workmates in the Department of Information Sciences at Mzuzu University for the moral and professional support during my PhD studies. Dr George Chipeta, Ms Sellina Kapondera, Mr Austin Phiri, Mr Fiskani Ngwira and Mr Luke Mwale all deserve my sincere appreciation. In a very special way, I thank Dr Chipeta for unselfishly sharing materials and for the many general fruitful engagements in research.

My sincere thanks go to UNI1 and UNI2 managements for granting me permission to conduct my research at these two institutions. I further thank all participants who participated in the study without being paid a coin or receiving any gift. You are wonderful people! Relatedly, I thank all research assistants who helped in data collection. I single out Mr Apatsa Selemani for doing extra-ordinary work in the data collection process at UNI2.

Family support cannot be ignored in such a long and terrifying journey. I am especially indebted to my fiancé, Joan Kacheche for her unflagging support through this journey. To my mother, Mrs Mary Mukumbwa, my only sister, Ms Emily Chawinga, my brothers, Acha, Moses, Chaupe and Michael and all the Chawingas, I say thank you all.

DEDICATION

I dedicate this research to my loving late father Boneck Mwasweshi Chawinga who inspired me at a very young age to excel in education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ABC	African Bible College
ALA	American Library Association
AIP	Archival Information Package
ANOVA	One-way Analysis of Variance
CCMF	Community Capability Model Framework
CCSDS	Consultative Committee for Space Data Systems
CHANCO	Chancellor College
CIF	Crystallographic Information File
CoM	College of Medicine
CUNIMA	Catholic University of Malawi
DCC	Digital Curation Centre
DIP	Dissemination Information Package
DMIF	Digital Microscopic Image Format
DOI	Digital Object Identifier
E-E-C	Education-Expertise-Curation
EU	European Union
Exif	Exchangeable Image File Format
FITS	Flexible Image Transport System
FAIR	Findability, Accessibility, Interoperability, and Reusability
ICPSR	Interuniversity Consortium for Political and Social Research
ICT(s)	Information and Communication Technologies
ISO	International Organization for Standardization
IT	Information Technology
JISC	Joint Information Systems Committee
KCN	Kamuzu College of Nursing
LIASA	Library and Information Association of South Africa
LIS	Library and Information Science
LUANAR	Lilongwe University of Agriculture and Natural Resources
MAGU	Malawi Assemblies of God University
MANCOSA	Management College of Southern Africa
MARC	Machine Readable Cataloguing
MARCXML	Machine Readable CataloguingeXtensible Markup Language

MAU	Malawi Adventist University
METS	Metadata Encoding and Transmission Standard
MGDS	Malawi Growth and Development Strategy
MODS	Metadata Object Description Schema
MSCE	Malawi School Certificate of Education
MUST	Malawi University of Science and Technology
Mzuni	Mzuzu University
NCHE	National Council for Higher Education
NCST	National Commission for Science and Technology
NRF	National Research Foundation
NUSL	National University Student Loan
OAIS	Open Archival Information System
OMB	Office of Management and Budget
ODL	Open and Distance Learning
PDF	Portable Document Format
PLU	Pentecostal Life University
PREMIS	Preservation Metadata Implementation Strategies
RDM	Research Data Management
SIP	Submission Information Package
SPSS	Statistical Package for the Social Sciences
UCT	University of Cape Town
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNIMA	University of Malawi
USA	United States of America
UWC	University of the Western Cape
XML	Extensible Markup Language
XMP	Extreme Memory Profiles

CHAPTER ONE

BACKGROUND TO THE STUDY

1.1. Introduction

Advancements in Information and Communication Technologies (ICTs) have created an environment where information in digital or electronic form is being produced at an unprecedented rate in various sectors and levels of the society. Large and small-scale businesses, industries, governments, universities, scientists, consumers and non-profit organisations are all generating data at an incredible pace (Gordon-Murnane, 2012). Public universities in Malawi have not been spared from the influx of various forms of digital information. This study set out to investigate a wide range of issues in regard to an emerging domain of Research Data Management (RDM) in Malawian public universities. The study focused on gaining a deeper and holistic understanding of how digital research data in Malawian universities is created, processed, organised, stored, preserved, disseminated and re-used. The investigation focused on three key stakeholders of the research process in higher education: librarians, lecturers (researchers) and directors of research who, according to Davidson, Jones, Molloy and Kejser (2014), have indisputably a clear-cut interest in RDM activities. Endorsing the preceding claim, Cox and Pinfield (2016) persuasively argue that libraries need to collaborate with researchers along with other key university players in pursuit of RDM activities.

1.2. From digital preservation to data curation and RDM

Until a few years ago, digital preservation was more concerned with superseding technological obsolescence. However, Higgins (2011, p. 79) reports that the focus has "shifted to ensuring that digital material is managed throughout its lifecycle so that it remains accessible to those who need to use it". The overarching aim is now to "ensure continued access to digital materials for as long as necessary" (Walters & Skinner, 2011, p. 5). To realise such long-term storage, preservation and access to digital data, the term digital curation was coined in 2001 (Kim, Warga & Moen, 2013, p. 94), initially as a seminar title on digital archives, libraries and eScience. According to Constantopoulos et al. (2009, p. 37), the central principle of digital curation is to ensure the future fitness of digital information which, as its context of use evolves, requires the active management and appraisal of digital assets over their entire lifecycle. Thus,

the literature seems to suggest that the concept of digital curation has become an umbrella term in the preservation of various forms of digital or electronic information.

Worth noting, however, is that the phrases digital curation and digital preservation are mostly used in a somewhat similar fashion. Walters and Skinner (2011) have laboured to make a slight distinction between the two terms, defining digital curation as the actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content. The two scholars define digital preservation as a series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Another term, which is increasingly being interchangeably used with digital curation, is data curation. The term has a similar meaning to that of digital curation. It has to be acknowledged, however, that despite authors choosing to use either the term data curation or digital curation, they all have the same goal of maintaining, preserving and adding value to digital data throughout its lifecycle.

Since there is no inherent distinction between data curation and digital curation, the choices in their usage have been attributed to discipline and cultural influences (Walters & Skinner, 2011; Zvyagintseva, 2015). On the cultural aspect, Zvyagintseva (2015) says the adoption of the term data curation by Yakel (2007), an influential American author on RDM, has influenced most Americans to prefer using this term whereas the popularisation of the term digital curation by the United Kingdom's (UK) Digital Curation Centre and others suggests a European preference. From the discipline perspective, Walters and Skinner (2011, p. 16) explain that data curation is mostly used in the fields of Science, Engineering and Social Science whereas digital curation is preferred by humanities and arts environments. In this study, the term data curation is adopted and this decision is influenced by the cultural aspect as highlighted by Zvyagintseva (2015). Most African literature related to this topic is dominated by the term data curation implying that the term is unarguably popular in Africa.

RDM is an extension of data curation and its definition is not contextual because a generic definition cuts across various discipline divides. In its generality, RDM is defined as the organisation of research data beginning from its entry into the research cycle, its sharing or dissemination, its storage, retrieval, re-use, security, and archiving its valuable results (Cox & Pinfield, 2016, p. 300; Whyte & Tedds, 2011, p. 1). In simple terms, Chiware and Mathe (2016, p. 2) refer to RDM as the "storage, access and preservation of data produced in particular investigations or research projects".

1.3. Researchers and the RDM era

It is common knowledge that lecturers are responsible for producing knowledge through various research studies that they conduct. In modern times, research has been profoundly boosted by the emergence of powerful computing technologies (Cox & Pinfield, 2016). Cox and Pinfield (2016) argue further that this pervasive use of ICTs in the research process has led to the generation and use of large data sets across disciplines. Universities have actually "made digital materials integral parts of their work" (Schumacher & VandeCreek, 2015, p. 97). In that regard, the emerging term of RDM is increasingly becoming a strategic priority for universities (Pryor, 2012; Whyte & Tedds, 2011). For instance, in the UK, United States of America (USA) and other developed countries, major research funders require applicants to indicate in their proposals how they are going to comply with RDM requirements for the preservation, sharing and reuse of research data in digital format (Cox & Pinfield, 2016; Heidorn, 2011, p. 663; Schumacher & VandeCreek, 2015, p. 97).

The literature shows that South Africa has set the pace for RDM activities in sub-Saharan Africa. For example, the National Research Foundation of South Africa has released a statement on open access to enforce the retention of research data for all research that it has funded (Chiware & Mathe, 2016). Moreover, Kahn, Higgs, Davidson and Jones (2014) indicate that the Library and Information Association of South Africa (LIASA) in conjunction with the UK's Digital Curation Centre hosted a workshop to equip library professionals in South Africa with RDM knowledge and skills. Considering that most African universities such as in Malawi are working towards situating themselves in the international research agenda and are mostly dependent on funding from the developed world, studies that aim to investigate aspects of research data management become imperative. The researcher aligns his line of thinking with that of Guedon (2015) who argues that researchers need to share their data with the understanding that new knowledge is built on existing knowledge hence the need to effectively process, preserve, share and re-use the newly created knowledge. Such developments are motivators of this heuristic study, which will investigate various aspects of RDM practices in Malawian public universities.

1.4. Librarians and the RDM era

The emergence of the concept of RDM has rightfully created new specialist roles in data curation (Digital Curation Centre, 2011; Hyams, 2008; Pryor & Donnelly, 2009; Swan & Brown, 2008). These roles include digital curator, data curator, and data manager (Kim et al.,

2013, p. 94; Swan & Brown, 2008). Walters and Skinner (2011) argue that in the data curation realm, there is a need for dedicated staff to prepare existing digital content for storage, access, the safe exchange between storage media and preserving it for long term management.

Some authors (Alvaro, Brooks, Ham, Poegel & Rosencrans, 2011; Corrall, 2012; Gabridge, 2009; Henty, 2008; Lyon, 2012; Monastersky, 2013) have suggested that librarians are better positioned to take a leading role in research data management. In that regard, it has been emphasised in the literature that librarians need to reinvent their roles by incorporating functions related to organising and manipulating data and data sets (Kim et al., 2013, p. 94). Already, librarians are involved in archiving and preserving data in universities through institutional repositories (Swan & Brown, 2008). It is plausible to agree with Kim et al. (2013, p. 94) that "the similar skill sets used in traditional library work may be beneficial to curation of work involving digital data and information". Given this context, investigating librarians' capabilities related to digital curation becomes justifiable, hence this study. As digital information epitomises all undertakings of academic libraries and information centres, interest has shifted towards understanding the new roles of librarians in this fast-paced and data-intensive information environment.

1.5. RDM: A converging zone for researchers and librarians

Considering that researchers are producers of knowledge which is commonly managed by librarians, it becomes reasonable to suggest that there is an inherent bond which ties researchers and librarians. Chiware and Mathe (2016) and Cox and Pinfield (2016) are of the view that the relationship between the library and researchers will be crucial in the implementation of RDM activities. Walters and Skinner (2011) argue that various groups involved in the research process are increasingly debating the best ways to preserve digital information for long periods of time. In sum, Constantopoulos et al. (2009, p. 37) conclude that data curation is an interdisciplinary domain that combines the skills and practices from many disciplines such as computer science, archival science, librarianship and information science.

1.6. Setting the context: Geographical, social and economic overview of Malawi

Malawi is a small landlocked country located in the southern part of Africa. It shares its international borders with three countries: Tanzania to the north and north east; Zambia to the west; and Mozambique to the South (National Statistical Office, 2017, p. 1). The previous national population census which was conducted in 2008 showed that the population of the country was 13,077 million (National Statistical Office, 2008, p. 3). Various recent reports

have independently confirmed that the population is rising at an alarming rate with the United Nations (2014) pegging it at 14.8 million in 2014 and three years later, that is in 2017, the National Statistical Office (2017, p. 2) pegged the population at 17.22 million. The population is projected to exceed 29 million by 2030 (United Nations, 2014). As of 2015, the country was positioned at 170 out of 188 countries and territories on the Human Development Index (IHDI) (United Nations Development Programme, 2016).

Economic activities in Malawi are diverse but the agricultural sector remains the major activity. The country's Gross Domestic Product (GDP) is US$3.5 billion and its Per Capita Income is US$290 (United Nation, 2014). The majority of the population (75 %) lives below the poverty line of less than US$1.25 per day (United Nations, 2014). The agricultural sector contributes over 82% of the country's total earnings and it further contributes 28.7 percent of GDP (United Nations, 2014) and more than 80 percent of export earnings (Kaluwa, 2010; United Nations, 2014). The agricultural sector is a source of over 80% of the national employment (Kaluwa, 2010).

1.7. An overview of higher education in Malawi and research output

The Malawi Government is very clear about its policy to improve access to quality education. This commitment is demonstrated in the Malawi Growth and Development Strategy (MGDS) that acknowledges education is "a catalyst for socio-economic development, industrial growth and an instrument for empowering the poor, the weak and voiceless" (MGDS, 2017, p. 50). The government commits to increasing its transition rates from primary to secondary school, and from secondary school to higher education. However, regardless of the commitment by the government to increase to basic and higher education, it appears it remains wishful thinking considering that the latest statistics symbolise a shocking limited number of students who proceed from primary to secondary school, and then to university. For instance, in 2017, only 16% of primary school students proceeded to secondary school while only 8% were admitted into various universities (MGDS, 2017, p. 50). It is therefore, not surprising that, for many years, the literacy rate has oscillated between 62% and 65% (MGDS, 2017, p. 38; United Nations Educational, Scientific and Cultural Organisation (UNESCO) Institute for Statistics, 2019), with the male literacy rate pegged at 73% while the female literacy rate is 59% (MGDS, 2017, p. 38). This literacy rate is low when compared to other Southern African countries such as Zimbabwe (89%) and South Africa (94%) (UNESCO Institute for Statistics, 2019). The poor literacy rates are a manifestation of the education system (primary, secondary and university) in Malawi.

In particular, it remains a far-fetched dream for higher education to absorb the many students who excel in the Malawi School Certificate of Education (MSCE) examinations which are a prerequisite for admission into any institution of higher learning in Malawi. This is because, generally, the higher education system in Malawi is poorly developed when compared to other countries in Sub-Saharan Africa. Financial constraint is the key factor crippling the operations of universities (Chivwara, 2013; Lombe, 2013, p. 275). While the main source of funding of public universities is the government, private universities operate entirely on fees from students (World Bank, 2010). For public universities, the fees are subsidised whereas for private universities, students pay exorbitant fees. The irony is that both, private and public universities are dominated by students from the Malawian wealthiest families (World Bank, 2010). Students from wealthy families study in premier high schools and they score better grades at national examinations or MSCE and stand better chances of being admitted into public universities whose admission rates are always low. Similarly, considering that private universities are too expensive for the poor, they are also dominated by students from wealthy families. Worse still, although the government has established the National University Student Loan (NUSL) to support needy students, only a limited number access such loans because there are no measures taken to discriminate against students from the wealthy families benefiting from this loan facility; the loans are eventually hijacked by the wealthiest thereby defeating the whole purpose of the loan facility. According to the World Bank (2010), the government loan facility is not extended to students in private universities implying that only students from elite families have access to such universities.

There are four public universities in Malawi (National Council for Higher Education (NCHE),2018a).They include the University of Malawi (UNIMA) which operates with five colleges that include Chancellor College (CHANCO), the Polytechnic, College of Medicine (CoM), and Kamuzu College of Nursing (KCN); Lilongwe University of Agriculture and Natural Resources (LUANAR), Mzuzu University (Mzuni) and Malawi Science of Technology (MUST). With increased demand for higher education, Malawi has witnessed an upsurge of private universities in the past two decades. In 2018, the NCHE had registered 16 private universities. However, of these 16 universities, the NCHE has accredited eight that includes the Catholic University of Malawi (CUNIMA), Pentecostal Life University (PLU), Malawi Assemblies of God University (MAGU), Malawi Adventist University (MAU), Management College of Southern Africa (MANCOSA), African Bible College (ABC), Daeyang University, and DMI St John the Baptist University (NCHE, 2018a). Most of these private universities are

run by churches and most of them were established after 1998 and are solely funded by the parent churches (Divala, 2013, p. 1).

Generally, public universities are grappling with running their core business of teaching and research because of the continued decrease in support from the government (Chivwara, 2013); most structures such as residences and classrooms have not been expanded to adequately accommodate the increasing demand of university education. This is despite unprecedented pressure these universities received from the same government to increase enrolments. In compliance with the government directive to increase enrolments, universities have opted to offer their programmes through Open and Distance Learning (ODL) (Chawinga & Zozie, 2016). The student population in the four public universities as of 2019 is presented in Table 1. To effectively produce highly qualified graduates, universities have recruited lecturers who are subject specialists in various disciplines. The total number of lecturers at each public university is presented in Table 1.

Table 1. Public universities in Malawi

University	Faculty	Student population	Staff population	Source
Mzuni	Education	4,500	150	Mzuni (2017)
	Environmental Sciences			
	Tourism & Hospitality Management			
	Humanities and Social Sciences			
	Science, Technology and Innovation			
CHANCO (UNIMA)	Social Science	5,000	221	CHANCO (2019)
	Science			
	Humanities			
	Education			
	Law			
CoM (UNIMA)	Biomedical Science and Health Professions	2 000	130	CoM (2017)
	Medicine			
	Public Health and Family Medicine			
KCN (UNIMA)	Nursing	500	80	KCN (2019)
	Midwifery			
Polytechnic (UNIMA)	Education and Media Studies	4,777	256	Polytechnic (2019)
	Built Environment			
	Applied Science			
	Commerce			
	Engineering			
LUANAR	Agriculture	5,738	150	LUANAR (2019)
	Development Studies			
	Food and Human Sciences			
	Natural Resources			
MUST	Academy of Medical Sciences	2,000	200	MUST (2019)
	Malawi Institute of Technology			
	Bingu School of Culture and Heritage			
	Ndata School of Climate and Earth Sciences			
Total		**24515**	**1212**	

A glance at research output in Malawian universities

According to Mitchell, Rose, and Asare (2018), 80-85% of research conducted in Africa is affiliated with institutions of higher learning and the largest outputs totalling 79% are peer-reviewed articles. Sub-Saharan Africa (SSA) contributes 0.72% of global research implying

that the region accounts for less than 1% of the world's total research output. Health sciences contribute more research output in SSA than any other discipline, accounting for 45% of all SSA research (World Bank & Elsevier, 2014, p.3).

In Malawi, the government recognises the value of research and in an effort to galvanise research output and to clear research bottlenecks that slow research activities, the government has established the National Commission for Science and Technology (NCST) that regulates research activities carried out by the various institutions and individuals in Malawi. The NCST has delegated some powers to the National Health Sciences Research Committee; and the College of Medicine Research and Ethics Committee, the two ethics committees that review research proposals to ensure that methodological and scientific rigor of research protocols are verified before their approval (Kirigia, Kathyola, Muula & Ota, 2015, p.5). The government has also developed and adopted two research agenda documents: the National Research Agenda in the Social Sciences and Humanities; and the National Health Research Agenda (Ministry of Health, 2012; NCST, 2013). The government has identified 14 key specific research priority areas as follows: social identities; physical and moral culture; communities on the margin; physical resources and infrastructure development and management; innovations, communications and technology culture; health and well-being; human resource development, management and utilisation; enterprise development, capital generation and financing; social and political transitions; peace and security; legal and justice systems; international relations; and education rural and urban farming systems (NCST , 2013).

Over the years, research output has been increased by Malawian researchers. UNESCO (2014, p.65) reports that statistics computed from Web of Science and SCOPUS which are the highly valued international databases, position Malawi at 107 in the world and 16 in Africa in terms of the production of peer reviewed articles. Considering that not all Malawian scientists publish their research articles in databases indexed by the Web of Science and SCOPUS, the number provided herein may not reflect a true picture of Malawi's research position at the continental and global level.. Nevertheless, the country's position is worth celebrating considering that it is placed in the band of poorest countries in the world. According to UNESCO (2014, p.65), the Malawi Government allocates less than 1% of its GDP to research and development. Malawian researchers need to be commended for publishing more in mainstream journals than other countries with a similar population and economic predicament. According to a report by the World Bank and Elsevier (2014, p.32), research output by various fields in Malawi is distributed as follows: Medical and Health Sciences (19%); Natural Sciences (15%); Engineering and

Technology (20%); Agricultural Sciences (17%); Social Sciences (18%); and Humanities (11%). A summary of research output and growth rates by field of science is provided by UNESCO (2014, p.55) as depicted in Table 2.

Table 2. Research output in Malawi by field of science (1967, 1977, 2007 & 2010)

Researchers per field of science [Head count]	1967		1977		2007		2010					
	Total	Percentage	Total	Percentage	Total	Percentage	Total	Percentage	Business	Government	Higher education	Privatenon- profit
Total researchers by field of science	**198**	**100**	**263**	**100**	**831**	**100**	**1843**	**100**	**n/a**	**507**	**1324**	**12**
Natural sciences	54	27	77	29	306	37	288	16	n/a	71	217	0
Engineering and technology	0	0	39	15	19	2	372	20	n/a	105	267	0
Medical sciences	0	0	4	2	157	19	343	19	n/a	93	242	8
Agricultural sciences	92	46	79	30	238	29	312	17	n/a	92	216	4
Social sciences	42	22	64	24	101	12	340	18	n/a	94	246	0
Humanities	0	0	0	0	10	1	188	10	n/a	52	136	0
Total female [F]	n/a	n/a	n/a	n/a	n/a	n/a	360	100	n/a	106	251	3
Natural sciences [F]	n/a	n/a	n/a	n/a	n/a	n/a	64	18	n/a	15	49	0
Engineering and technology [F]	n/a	n/a	n/a	n/a	n/a	n/a	24	7	n/a	12	12	0
Medical sciences [F]	n/a	n/a	n/a	n/a	n/a	n/a	60	17	n/a	17	43	0
Agricultural sciences [F]	n/a	n/a	n/a	n/a	n/a	n/a	39	11	n/a	19	17	3
Social sciences [F]	n/a	n/a	n/a	n/a	n/a	n/a	94	26	n/a	27	67	0
Humanities [F]	n/a	n/a	n/a	n/a	n/a	n/a	79	22	n/a	16	63	0

Source. UNESCO (2014, p.55).

1.8. Research problem

The Malawi Government acknowledges that research has a role to play in fostering the socio-economic development of the country. Internationally, the Malawi government draws her recognition of the importance of research from various international instruments and groupings such as those developed by the UNESCO to which Malawi is a signatory thereby embracing its declarations, protocols and conventions (NCST, 2013a). Nationally, the Government's commitment for innovative research is inspired by the MGDS (2017) and the Vision 2020 (2000) which spell out national development priority areas whose successful execution will

depend on innovative research by researchers in various disciplines such as the social sciences and humanities and clinical research. To accelerate research outputs at national level, the Malawi Government has developed two research agenda documents: the national research agenda in the social sciences and humanities and the national health research agenda (Ministry of Health, 2012; National Commission for Science and Technology, 2013) which are used as the guiding documents for researchers in various universities.

According to UNESCO (2017), research outputs in Malawian universities have doubled over the years and most of the researchers are publishing in open access journals. For example, 156 researchers have published their research articles in BioMed Central and 186 articles have been published in Public Library of Science (PLOS) open access journals such as PLOS ONE, PLOS Medicine and PLOS Neglected Tropical Diseases (UNESCO, 2017). Local open access journals have been established in which Malawian researchers are publishing their research outputs. Publishing in international and local open access journals has significantly increased citations thereby strengthening the ranking of Malawi in respect of clinical medicine, immunology, microbiology, agricultural and social sciences. Many more university researchers are conducting research studies whose findings are being published in non-open access journals.

An increase in research activities coupled with the use of computer software in research activities has in turn led to the production of large amounts of digital research data. As already noted earlier, like other researchers across the globe, researchers in Malawian universities are sharing their findings by publishing in both, open access and subscription journals. The fundamental question that needed to be answered is *what do these researchers do with research data after using it for writing unpublished reports, journal articles and conference presentations?* Moreover, since digital data can easily be lost or corrupted (Consultative Committee for Space Data Systems, 2002, p. 5; Cox & Pinfield, 2016), what data preservation practices have researchers in Malawian public universities adopted to safeguard their research data against loss due to software and hardware failure? To avoid giving hypothetical answers to the preceding question, this study set out to investigate various issues about research data management practices amongst researchers in Malawian public universities.

Moreover, a systematic search in major libraries and online databases such as Google Scholar, EBSCO and Scopus returned no results about research data management in a Malawian context. This underscored a gap in this area of research in Malawi, hence one of the compelling

justifications for this study. Furthermore, prior studies that have explored research data management from the combined perspectives of researchers and librarians are seemingly scarce in the literature. The present study was spurred on by these issues to investigate research data management in two Malawian public universities. This is important considering that many standards and best practices for data curation advocate for institutions to manage selected materials in archives, repositories and data centres (Higgins, 2008).

1.8.1. Aim and objectives of the study

This aim of the study was to investigate research data management practices in public universities in Malawi focusing on how research data is generated, organised, shared, stored and preserved for the purpose of re-use and long-term access. The following research objectives were formulated to help address the research problem:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers need to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affect the management of research data in public universities in Malawi

1.9. Significance of the study

An increase in research activities coupled with the use of computer software in research activities has in turn led to the production of large amounts of digital research data. It is however argued that data sets are potentially fragile, vulnerable to storage failures and technological obsolescence (Beaudoin, 2011; Consultative Committee for Space Data Systems, 2002, p. 5; Cox & Pinfield, 2016; Pogue, 2009; Schumacher & VandeCreek, 2015; Shen, 2016). This study is therefore timely because, based on its findings, it has made recommendations to public universities about how to manage research data for posterity, longevity and re-use. The study has investigated and demonstrated the disjointed and obscurely explained (in the literature) relationship between lecturers, (researchers) and librarians in executing data curation activities thereby adding important scholarly literature to this emerging discipline. After all, Chen and Wu (2017, p. 346) argue, that to make sure that libraries can provide personalised, specific and effective services for researchers, it is necessary to

understand the current situation of research data management and needs for RDM services in the research process.

1.10. Scope and limitations of the study

The study's focus was on research data management in public universities in Malawi. Specifically, two public universities, UNI1 and UNI2 were included in the study. Other public universities (MUST and LUANAR) and all private universities were not included in the study. In terms of research participants, the study targeted library staff, lecturers and directors of research in the two universities.

The key limitation of the study was financial constraints. Although it was by design to leave out private universities, it was not by choice to exclude the other two public universities; the researcher did not have adequate funding to extend this study to other public universities.

1.11. Definition of key terms

Definitions of some key terms as used in the context of this study follow. Considering that the subject area under study is broad and full of important terms, only terms fundamental to the study are defined.

Research data

These are "the factual records (e.g. microarray, numerical and textual records, images and sounds, etc.) used as primary sources for research, and that are commonly accepted in the research community as necessary to validate research findings" (CARL Data Management Sub-Committee, 2009, p. 4).

Research data management

The organisation of data beginning from its entry into the research cycle, its sharing or dissemination and its storage, retrieval, reuse, security and or archiving its valuable results (Cox & Pinfield, 2016, p. 300; Whyte & Tedds, 2011, p. 1).

Digital curation

Digital curation refers to the process of maintaining, preserving and adding value to digital research data throughout its lifecycle (Digital Curation Centre, 2017).

Data curation

The term data curation is used interchangeably with the term digital curation in the literature. Like digital curation, data curation is defined as "The active and on-going management of data "through its entire lifecycle of interest and usefulness to scholarship, science, and education" (Noonan & Chute, 2014, p. 203).

Data deluge

The term "data deluge" was coined in the early 2000s (Hey & Trefethen, 2003) to reflect the sheer volume and magnitude of research data in the digital age.

1.12. Outline of the thesis

The thesis has been organised into seven chapters as described below.

Chapter One: Background to the study: This chapter introduced the topic of the research highlighting ongoing debates concerning data curation. In particular, the chapter offered insights about the interconnectedness between research data management and researchers and librarians. The chapter also defined the research problem that influences the researcher to carry out the study.

Chapter Two: Review of literature: This chapter covers a comprehensive review of the relevant empirical and theoretical literature in both print and electronic format using the research objectives as a basis. The chapter identifies the topics related to the study which are dominating this area of research. The chapter puts into perspective the role of researchers and librarians in the research data management process. The chapter identifies the gaps in research that are in turn addressed in this current study.

Chapter Three: Conceptual frameworks: This chapter provides a comprehensive review of the conceptual frameworks including but not limited to Community Capability Model Framework (CCMF) and Digital Curation Centre Lifecycle Model. The chapter justifies the reasons why doctoral research should be guided by frameworks and models. After reviewing various models in this chapter, two of them are adopted to guide the current study.

Chapter Four: Research methodology: This chapter presents the research methodology covering the pragmatist paradigm, qualitative and quantitative approaches, research design and survey research designs, population of study, sampling procedures, instrument validity and reliability, data collection, data analysis and ethical issues of research.

Chapter Five: Data analysis and presentation of findings: This chapter presents and analyses the findings of the study based on the specific questions of the study, literature and theoretical constructs informing the study. Verbatim quotes for qualitative findings, inferential and descriptive statistics are used to present the findings.

Chapter Six: Discussion of findings: This chapter discusses the findings by contextualising them in the extant literature and the adopted theories. The chapter checks and ensures that the aim and objectives of the study are met.

Chapter Seven: Summary of findings, conclusion and recommendations: This chapter presents a summary of the findings, conclusions and recommendations of study. The originality and contribution of the study is presented in this chapter. Areas for further research are also suggested in this chapter.

CHAPTER TWO

LITERATURE REVIEW

2.1. Introduction

Card (2012, p. 727) defines literature reviews as "systematic syntheses of previous work around a particular topic". It involves the researcher consulting various but vital information sources such as academic texts, review articles, reference databases and public data (Winchester & Salji, 2016, p. 308). Wellington, Bathmaker, Hunt, McCulloch, and Sikes (2005) caution that masters and PhD students should not have a misconception that by reviewing the literature as a separate section means this academic activity ends here. The literature review is an ongoing activity and it is part of the entire research process which spans from the first chapter to the final one. Doctoral students should "not stop reading before you have submitted your thesis (keep reading until the last minute)" (Wellington et al., 2005, p. 16).

2.2. Establishing the need for a literature review in a doctoral thesis

Before delving into other aspects of the literature review, it becomes important to highlight the reasons why researchers need to conduct this important academic exercise. Literature reviews are critical for any scholarly work because, apart from being a crucial appraisal for a subject of interest, it is an academic requirement necessary for research planning and contextualising the findings (Winchester & Salji, 2016, p. 308). It will enable the researcher to identify prior research that supports or differs from findings of current research thereby enabling the researcher to situate the research in the field (Winchester & Salji, 2016, p. 308). It aims to establish what research has been done in the field of study while enabling the researcher to identify the gap or the further contribution the study will make to the field (Wellington et al., 2005). Through a literature review, the researcher identifies the common theories and conceptual frameworks in the subject area, identifies methods and approaches that are commonly used to investigate the field of study (Torraco, 2016, p. 405; Wellington et al., 2005).

Most scholars place an emphasis on identifying the gap as one of the most important outcomes of undertaking the literature review. The literature is not about listing an exhaustive list of all that has been published in a related field; rather, it has to be informative and personal but unbiased summation that provides both supporting and conflicting findings, inconsistencies and viewpoints (Winchester & Salji, 2016, p. 309). Most scholars agree that conducting a

literature review helps the researcher to consolidate what is already known and in turn, identify any knowledge gaps that new research could focus on (Winchester & Salji, 2016, p. 308). The knowledge gap identified in the literature is what Card (2012, p. 727) calls "unresolved questions or next steps for future research".

2.3. Demarcation of the chapter

There seems to be no universally accepted standard for organising literature. Most authors leave the choice to the researcher. The researcher may organise the literature by history and methodology (Card, 2012; Torraco, 2016, p. 405). However, the most recommended procedure is by arranging the literature by research questions or objectives (Card, 2012; Wellington et al., 2005; Winchester & Salji 2016, p. 308). This chapter is divided into six sections as discussed below.

The first section clarifies some terms commonly related to the concept of data curation. These terms include data, digital preservation, and e-science, and provides a clarification for the difference between data curation and digital curation. The section concludes by discussing why the term data curation has been adopted in this study.

The second section focusses on understanding various issues in regard to the involvement of librarians in RDM activities. The broader themes covered include the role of librarians in the data curation process, RDM competencies for librarians and RDM services offered by librarians.

The third section seeks to review various aspects in relation to data management practices by researchers. The section discusses two broad topics: research data creation and storage practices amongst researchers; and research data sharing, preservation and re-use amongst researchers.

The fourth section reviews literature on cyberinfrastructure. Three key issues are discussed in this section and they include software, storage facilities, and metadata standards and interoperability.

The fifth section is about factors that affect the development and implementation of data curation activities. Issues discussed include costs of curatorial activities; lack of RDM expertise; restrictive institutional policies; rights management issues; obsolescence of technologies; ethical and legal norms; incentives for researchers; and RDM terminological differences.

The final section reviews two evidence based interventions in data curation with a focus on the UK's DCC and the de.NBI Systems Biology Service Centre in Germany.

2.4. Data curation: Analysis of some related concepts

The emergence of digital environments in the research realm has led to the creation of many forms and complex born digital information resources that continue to proliferate on a daily basis. Across these various disciplinary divides, scholars produce a myriad digital scholarly works including "scientific data, notes, electronic records, arts and new media, multimedia learning objects, user-generated web content, and the products of mass digitization efforts" (Walters & Skinner, 2011, p. 11).The accelerated pace by which research institutions are generating and acquiring digital information has led to the birth and rise of various terms and concepts that aim at creating, managing, preserving and offering access to these digitally produced objects. There are various terms associated with the preservation of digital content or materials. This section therefore, discusses some of the popular terms while highlighting their intersections with data curation where necessary. These terms include research data, e-research, and digital preservation. Some of these terms were reviewed in Chapter One; therefore, this sections aims at providing a deeper analysis.

2.4.1. Research Data

The term research data is not new as it is at the centre of the whole research process. Research data refers to various forms of factual records that are used as sources for primary research; examples of such records include microarray, numerical and textual records, images, and sounds (CARL Data Management Sub-Committee, 2009, p. 4). According to the CARL Data Management Sub-Committee (2009, p. 4), these records are commonly accepted in the research community as necessary to validate research findings. Because data types can vary extensively in different disciplines and institutions (Krier & Strasser, 2014), it has been defined and conceptualised according to those disciplines and institutions (Ohaji, 2016, p. 25). Data can be categorised in three key ways. It can be categorised based on the processes used to gather or generate data e.g. experimental, computational/simulation data; derived data; data storage solutions; and data curation (National Science Board, 2005, p. 18; Thomas, 2011, p. 38). Other authors (Krier & Strasser, 2014) have categorised data based on forms (i.e. qualitative and quantitative data) and the level of gathering it (primary and secondary data) which could be generated through observational, experimental or computational methods (National Science Board, 2005, p. 18; Thomas, 2011, p. 38). Finally, data can also be categorised based on the

main categories of the subject areas. For instance, Borgman (2015) argues that, whereas in sciences and social sciences the distinction is based on 'raw' and 'processed' data, in humanities, the distinction focusses on primary, secondary and tertiary data. To avoid limiting the scope of data types, the present study embraces all three data types as described by scholars in this section.

2.4.2. Digital preservation

Walters and Skinner (2011) refer to digital preservation as activities that are taken to ensure continued access to digital content for the rest of its life span. Similarly, digital preservation is "an archiving activity in which specific items of data are maintained over time so that they can still be accessed and understood through successive change and obsolescence of technologies" (Yakel, 2007, p. 338). The understanding is therefore, that digital content can be managed and accessed as long as it remains valuable to the users or the generations to come. Preservation aims at ensuring that items or collections remain accessible and viable in subsequent technology environments. For instance, the University of Alberta Libraries aim to preserve their wealth of knowledge in digital format for the next 500 years (Zvyagintseva, 2015) regardless of technological changes and obsolescence that are vital for access to digital content.

2.4.3. E-science

Some researchers have paid particular interest in the term e-science which is commonly associated with RDM. According to Hey and Hey (2006, p. 517), e-science, commonly called e-research, has come to mean a set of tools and ICTs that are adopted to facilitate and offer support for collaborative and networked research. "Given the capabilities of cyberinfrastructure, collaborative and networked research is done within and across disciplines with much data being generated" (Ohaji, 2016, p. 25). The main objective of e-science is to enable researchers to do their research in faster and efficient ways using technological applications that enable them to access, manipulate and mine data (Hey & Hey, 2006, p. 517). E-research has been characterised by Lewis (2010, p. 8) as data intensive whereby researchers generate and use large volume of data through collaboration which affords researchers across multiple institutions to work together, share and use data using various available technologies and networks. The key advantage that e-science has brought to the scholarly community is that it has profoundly reduced the barriers of time and geography to collaborative research thereby leading to the production of valuable and large quantities of research data (Carlson, Fosmire, Miller & Nelson, 2011). A study that focused on finding out the reasons researchers are

involved in collaboration was conducted by Day (2008) who found that collaboration plays a key role in the data curation process by pooling resources together.

2.4.4. Data preservation, data curation or digital curation: What is the right term?

Data preservation refers to management practices that lead to the long term preservation of data. The concept of data curation is basically an extension of data preservation. According to Charbonneau (2013) and the Digital Curation Centre (2011), the data curation process is a digital curation process which aims to add value to digital research data so that it is well maintained, preserved, accessed and re-used throughout its lifecycle. Data curation aims to organise, display and repurpose preserved data collections. However, with the term data curation increasingly being used to encompass all activities of managing research data from creation to long term preservation, the term data preservation has become blurred with RDM activities. Data curation is thus concerned with addressing challenges of managing data produced as a result of research through planning, selection, preservation, description, management, edition, and reuse of data over time (Zvyagintseva, 2015, p. 1).

The term digital curation was first coined in 2001 as a title for a workshop which drew participants from various disciplines that met to map solutions to urgent challenges confronting "the long-term management of, and preservation of access to, digital information" (Kim et al., 2013, p. 67). According to Kim et al. (2013, p. 67) and Walters and Skinner (2011), digital curation are the activities that individuals perform to add value to digital content over its lifecycle. Activities practised in digital curation include those taken at the creation of the digital content (Walters & Skinner, 2011). Worth mentioning is that "the terminology for digital curation is not yet stable" (Kim et al., 2013, p. 77). This is perhaps the reason why various researchers have come up with various but similar definitions as presented in Table 3.

Table 3. Definitions of data curation

Sources	Definitions	Key points
Digital Curation Centre (2017)	It involves maintaining, preserving and adding value to digital research data throughout its lifecycle	The focus is on achieving longevity of digital research data
Walters and Skinner (2011, p. 5)	The actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content	Managing data starts at the data creation stage and the process is continued as long as the data remains valuable.
Zvyagintseva (2015, p. 4)	The practice that addresses the challenges of maintaining digital information over its entire lifetime, as long as it is useful to researchers	Access, dissemination, and preservation of both information content and context –data and metadata
Noonan and Chute (2014, p. 203)	The active and on-going management of data "through its entire lifecycle of interest and usefulness to scholarship, science, and education".	Managing data to ensure it is fit for contemporary use and available for discovery and reuse
Rusbridge et al. (2005)	Curation is the active management and care of data	Preservation and maintenance of data so that access is guaranteed when needed.
Permanent Access to the Records in Europe (2009)	The careful storage of all research output in such a way that it remains accessible, usable and understandable over the long term.	Data preservation for access, uniform standards and re-use

From the definitions presented in Table 3, it can be acknowledged that although authors choose to use either the term *data curation* or *digital curation*, they all point to a common goal of maintaining, preserving and adding value to digital research data throughout its lifecycle. This suggests that the terms *digital curation* and *data curation* can be used interchangeably without creating any confusion. As noted in Chapter One, use of either digital curation or data curation is rooted in cultural and discipline connotations. In terms of culture, data curation is popular in the USA while digital curation is a preferred terminology in Europe (Zvyagintseva, 2015). In relation to discipline, data curation is preferred in Science, Engineering, Social Science and allied fields whereas digital curation is popular in the humanities and arts environments (Walters & Skinner, 201, p. 16).

From the definitions presented in Table 3 together with that of Digital Curation Centre (2017), it is appropriate to conclude that data curation is about activities that are involved in managing and promoting the use of data from the time it is created to ensure that it remains suitable for contemporary use, and that it is easily accessible and retrievable for re-use or repurposing for ever. Observing that there is no inherent differences between data curation and digital curation, the present study uses the terms interchangeably.

2.5. Factors driving the popularisation and adoption of data curation

The literature shows that there are a number of factors that have contributed to the popularisation and adoption of the data curation concept across the globe. This section discusses some of these key factors.

2.5.1. Data deluge

Previously, research data was considered a by-product of research activities and research papers or reports were considered the main outcomes. However, in recent years, it has been proven that research data is a commodity worth managing and preserving (Davenport & Patil, 2012; Matlatse, 2016). Increasingly, a paradigm shift is being embraced by various scholars, research institutions and researcher stakeholders about the value of research data. Debates about data sharing and re-use are not new. They emerged in the 1980s when researchers became more concerned about the contribution that data sharing and re-use could make to the advancement of scientific research (Fienberg, Martin, & Straf, 1985; Glaeser, 1990) and attention to these issues has grown steadily (Yoon, 2015). However, the terms data sharing, data re-use and data curation have become significant and popular in the last 10 years following the emergence of the new research practice known as data-intensive research or e-science (Kunze et al., 2011). Within the broader domain of digital curation, there has been a rise in the need to provide long-term preservation, access and re-use of research data especially in data intensive science which is characterised by the emerging problems of data deluge (Kahn et al., 2014; Kim et al., 2013, p. 67; Schumacher & VandeCreek, 2015, p. 97; Walters & Skinner, 2011). Data deluge is a term that literally means *flood of scientific data* (Hey & Trefethen, 2003) meaning that in the era of e-science, there is unprecedented creation of data which needs to be well managed, preserved and reused.

2.5.2. Funding compliance

Both scholarly and professional literature identify funding compliance as one of the primary driving factors compelling researchers to embrace and engage in RDM activities. Generally, while all researchers may find it a good practice to manage their data through structured or unstructured procedures, funders are more interested in data sharing and data management plans (Charbonneau, 2013; Chen & Wu, 2017, p. 346). These open data policies orchestrated by funders are effectively promoting research data sharing and reusing during the research data lifecycle (Brambilla, 2015; Chen & Wu, 2017, p. 346; Enke et al., 2012; Wallis, Rolando & Borgman 2013). Kim et al. (2013, p. 67) report that in the USA a circular, issued through the

Office of Management and Budget (OMB), compels all federally funded researchers to make research data accessible to the public in various ways. Most other funding data policies require all grant recipients, in addition to making their research outputs such as journal articles publicly accessible, make their research data publicly accessible as well (Charbonneau, 2013; Schumacher & VandeCreek, 2015, p. 97). Similarly, in Europe, the European Union (EU) which is one of the major research funders across the globe takes RDM issues very seriously. To underscore its commitment towards data sharing across the EU region, the EU has dictated that from 2014 all data produced with EU funding should be accessible for free (European Commission, 2012).

Kim et al. (2013, p. 67) observe that the development and implementation of RDM requirements are underway around the world. South Africa has set the pace in Africa. Chiware and Mathe (2016, p. 2) , Koopman (2015), and Matlatse (2016) report that the National Research Foundation, which is a government agency responsible for all research activities at national level, has released a statement putting it as a condition that researchers who receive its funding are required to deposit the findings in open access repositories. The statement further compels researchers to deposit the data generated through such research activities in accredited open access data repositories. In view of these developments, Koopman (2015) predicts that sooner or later, RDM may become mandatory for South African researchers.

Researchers are therefore, increasingly succumbing to the demands of funders because they are aware that failure to adhere to data management requirements demanded by funders, will deny them access to scarce and treasured grant opportunities. In their international study, Huang et al. (2012) examined the attitudes, experiences, and expectations of biodiversity researchers regarding data sharing and archiving. The study revealed that, while 60 % of the researchers were willing to share their data, the only compelling factor was data sharing policies adopted and enforced by funders and publishers (Huang et al., 2012). Despite a strong case being made in the literature that funding agencies are key motivators for RDM, Schumacher and VandeCreek (2015) report of an additional interesting motivating factor which gives hope to the sustainability of RDM practices. Schumacher and VandeCreek (2015, p. 106) report that digital preservation measures at individual and institutional level do exist and are not bound by a grant body's needs. This is good news considering that not all research projects are funded and more so, not all funders may demand researchers deposit their data in open access data repositories for free distribution.

2.5.3. Journal policies

Journal policies and reviewers are also contributing to the enforcement of good data management practices. Most reviewers are increasingly demanding researchers submit their manuscripts alongside the underlying data. Reviewers asking for data to be submitted alongside manuscripts for review purposes cannot be questioned if assertions by Pitt and Tang (2013 are to be taken seriously by the scholarly community. "These are often the types of questions that can arise during the review of a manuscript, where curious reviewers might ask that additional analyses be included prior to publication" (Pitt & Tang, 2013, p. 216).This is necessary but also controversial (Koopman, 2015, p. 14). It is a necessity because it helps to ascertain the originality of the research and to deter fraud in research thereby achieving robustness of research findings but controversial because some quarters in the research community believe this may lead to theft of research data unless clear policies and data ethics are put in place by publishers (Doorn, Dillo & Van Horik, 2013). Some popular science journals such as *Atmospheric Chemistry and Physics, F1000Research, Nature,* or *PLOS One* have adopted data sharing policies with the objective of promoting public access to research data (Fecher, Friesike, & Hebing, 2015, p. 1). A bibliometric study by Piwowar (2011) examined how frequently researchers openly archived raw gene expression microarray data. From the 11,603 articles published between 2000 and 2009, it was revealed that the researchers were more likely to share the data if their findings were published in journals with strong data sharing policies.

2.5.4. Open science movement

Recently, movements that advocate for open access to research and data have emerged and are increasingly gaining momentum. Their primary aim is to advocate the sharing of data and greater experimental transparency (Mundel, 2014; Shen, 2016). These movements have innovatively won the support of major global donors such as the World Health Organization, the National Institutes of Health, the Wellcome Trust, the Research Councils UK and the Gates Foundation which have all demonstrated commitment to sharing research data and information (Shen, 2016). Open access movements have one key priority: make research freely available and allow the collection and sharing of data so that other scientists and health experts can access the latest evidence, draw on it to advance their own research, and benefit from this knowledge (Mundel, 2014). Guedon (2015) makes a good point worth highlighting that for many centuries, researchers have learned to share their papers or research findings, now they must learn to share their data. In his view, sharing the interpretation of data in the form of published papers or

unpublished reports is simply to optimise the whole research process hence, the need to share data as well is the very essence of science if science is to be conceived as a gigantic system of distributed intelligence.

The review in this section is proficiently summarised by Wicherts and Bakker (2012, p. 74) who attest that data sharing by scholars is being enforced by a compendium of reasons that include

> *...abiding by the scientific principle of openness, keeping the data for posterity, increasing one's impact, facilitation of secondary analyses and collaborations, prevention and correction of errors, and meeting funding agencies' increasingly stringent stipulations concerning the dissemination of data.*

2.6. Data curation: The emerging discipline for librarians

The proliferation of the digital environment has inevitably transformed the roles of librarians. To date, the literature is replete with evidence that librarians have successfully played an important role in e-publishing with the focus on open access publishing, e-print publishing and editing digital humanities or social sciences resources (Walters & Skinner, 2011, p. 6). Librarians have an opportunity to reposition themselves in the digital curation landscape by proactively engaging with researchers in their research activities, starting from the conceptual stages. Librarians in research intensive universities are responsible for hosting digital content through their institutional repositories. According to Walters and Skinner (2011, p. 17), librarians are responsible for hosting a broad range of digital content including digitised collections, licensed content, web archiving, research data, e-prints (research publications and electronic theses and dissertations) and digital instructional materials (digitally captured lecture series, symposia, and other campus events). By managing their institutional repositories and digital libraries which are very popular in modern librarianship and the academic world implies that librarians already have skills and knowledge which they can apply in managing research data. As noted already, data curation is just one of the aspects of digital preservation. It is therefore not surprising that Matlatse (2016) and Walters and Skinner (2011, p. 17) and report that librarians are increasingly responding to the needs of digital research data management.

2.6.1. Roles of librarians in the data curation process

Charbonneau (2013), Matlatse (2016), and Walters and Skinner (2011, p. 17) highlight various emerging roles for librarians in the digital preservation ecosystem which cut across various

disciplines such as the sciences, engineering and social sciences. Librarians are directly or indirectly involved in the creation of research data and information objects in which they proactively carry out data modelling, managing and capturing any content that comes from research teams. Another role involves managing research data. In this role, librarians are challenged to collect, ingest, describe, perform provenance-tracking (the information that documents the history of the data), provide access and re-use, integrate and preserve data (Charbonneau, 2013; Walters & Skinner, 2011, p. 17). A further different role is more about RDM in cyberspace which aims to offer collaboration in virtual communities. Here, Walters and Skinner (2011, p. 17) challenge librarians to use web-based tools to bring researchers with similar research interests together so that they participate in particular research activities virtually. Generally, researchers have poor data management skills and librarians can take this as an opportunity by offering regular and demand driven training to improve researchers' skills (Charbonneau, 2013). Librarians may also in some cases help researchers in creating, preparing and implementing data management plans as required by research funders (Charbonneau, 2013; Walters & Skinner, 2011). Numerous researchers have cited Heidorn (2011, p. 663) about the reasons why librarians seem to have a crucial role to play in curating research data:

- Curation of data is within the libraries' missions, and libraries are among the only institutions with the capacity to curate many data types;
- The data is critical to the scientific and economic development of society;
- There is a large volume of data not currently being curated adequately; and
- Governmental and non-governmental funding bodies are beginning to recognise the importance of data and are creating rules for people receiving funds for research and development.

As already observed earlier that more and more funding agencies require researchers to submit research data sharing plans, Charbonneau (2013) advocates that librarians can play a key role by providing guidance for meeting with funding requirements. Noting that researchers are more interested in day-to-day scientific research than in RDM activities (Shakeri, 2013, p. 73), librarians may exploit this situation to become ambassadors of RDM.

2.6.2. RDM competencies for librarians

During the past few years, issues relating to research data management in general and libraries' data management services for researchers in particular have attracted a great deal of attention in the library community (Chen & Wu, 2017, p. 346). There is a growing demand that librarians

need to acquire new types of skills and competencies in order to assume the new roles of digital curation (Heidorn, 2011; Newton, Miller & Bracke, 2011; Ray, 2012). In light of the emergence, popularisation and adoption of digital curation in research based universities, librarians changing roles and their new competencies have been explored in the scholarly and professional literature (Kahn et al., 2014; Kim et al., 2013; Newton et al., 2011; Walters & Skinner, 2011). Empirical studies have been conducted to re-examine librarians' roles in this so called emerging discipline for librarians.

Kim, Warga and Moen (2013) analysed job advertisements in the field of digital curation as posted in key online forums such as the American Library Association's JobLIST, Association of Research Libraries' Job Announcements, the Special Libraries Association's Career Centre Library and Information Science (LIS) Jobs and Digital Curation Exchange. The key knowledge, skills and abilities mostly sought by half of the employers were the ability to work in the ICT intensive environment. In particular, knowledge of multiple operating systems (UNIX/Linux); programming and scripting languages (Java, PHP, and Perl); HTML and other Web-related mark-up languages; relational databases (MySQL and Oracle) and advanced graphics software were highly sought after. Other attributes needed by employers included familiarity with and knowledge of various metadata standards, such as Machine-Readable Cataloging (MARC), Dublin Core, Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), and Preservation Metadata: Implementation Strategies (PREMIS) (Kim et al., 2013, p. 74). The fact that the these results show further that 132 (76%) employers expected applicants to perform digital curation activities signals that advanced ICT skills need to be embedded in Library and Information Science (LIS) schools' curricula. Charbonneau (2013) emphasises that librarians need to master advanced specialised skills in data analytics, visualization, relational databases and data mining. Although Kim et al.'s (2013) study indicates that 130 (75%) preferred applicants with master's degrees from ALA accredited institutions, it fails to show if most or all ALA accredited LIS schools offer the required advanced ICT skills demanded by the employers. It is however, well documented that iSchools have rich components of ICT courses; the subjects include social informatics, data management, information architecture, and digital libraries (Nalumaga, 2017). The iSchool movement has been embraced by the East African School of Librarianship at Makerere University, the only iSchool in Africa. Generally, LIS schools in Africa continue to assimilate ICT courses into their curriculum (Nalumaga, 2017; Raju, 2013) though they lag behind when compared to certified iSchools (Nalumaga, 2017).

If employers in Kim, et al.'s (2013) study were based in Africa, the employers could surely have engaged computer science or information technology graduates because the literature shows that it is rare to find LIS schools in Africa offering advanced ICT related courses. This view is supported by Kahn et al. (2014, p. 302) who report that in South Africa for example, most librarians are of the view that librarians are less competent in ICT skills for RDM and have a perception that ICT professionals are better positioned to carry out RDM activities. However, it is not too late for librarians to formally or informally acquire the knowledge and core competencies because Matlatse (2016) reports that the concept of RDM is still in infancy in Africa. In that regard, these librarians have a chance to enrol in educational programmes and training that can fully equip them with such curatorial skills and knowledge (Charbonneau, 2013, p. 366; Heidorn, 2011; Matlatse, 2016; Ogburn, 2010).

A master's study by Matlatse (2016) focused on investigating if training workshops on RDM could enhance RDM the knowledge and skills of librarians working in universities and research institutions in South Africa. The study identified RDM services that librarians can offer to researchers and they include data management (metadata administration, data preservation and archiving), crafting data management plans, facilitating access to research data and offering training to researchers (Matlatse, 2016, p. 81). In order to effectively offer these services, librarians need to amass competent skills in metadata, data referencing and citation, documenting data, data storage and security, data licensing, data management planning, managing repositories and formulating RDM user training guides (Matlatse, 2016, p. 81). The study found that although such training workshops played a key role in librarians' understanding and knowledge of RDM concepts, the workshops failed to instil core RDM skills and competencies in librarians. In the light of the findings, Matlatse (2016) emphasised the importance of formal education and recommended that LIS professionals should enrol with universities that offer courses in RDM. The good news however, is that, in Africa, librarians may study RDM at the University of Cape Town (UCT) which is offering a masters course in data curation. The study also noted that online workshops could be ideal sources for RDM skills for librarians.

Since digital curation is a new or emerging concept (Higgins, 2011; Matlatse, 2016) that requires new skills, Borgman (2010) recommends that librarians' skills and expertise in this domain be adapted through partnerships and formal education in order to effectively manage research data. These claims are reinforced by Ohaji's (2016) study in Australia which showed that there was a need to train librarians or re-skill or up-skill their roles in RDM. Almost all

staff involved with RDM will need training (Brown, et al., 2015). Table 4 summarises some core RDM skills that librarians need to acquire as synthesised from the literature.

Table 4. RDM training needs/skills for librarians

Dimension	Training needs/skills	Sources
Research	The need to understand various aspects of research such as research cycles, research project management and e-Research	Heidorn (2011), Newton et al. (2011), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Technology	The need to understand the various technologies, available research tools and the ever-changing ICTs, database development and software skills	Ng'eno (2018), Kahn et al. (2014), and Ohaji (2016).
Information management	The need to understand information governance and access principles (international or existing ones), informatics and interchange standards	Heidorn (2011), Newton et al. (2011), Ng'eno (2018), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Research data	The need to understand data collection, research data curation and management	Heidorn (2011), Kahn et al. (2014), Newton et al. (2011), Ng'eno (2018), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Metadata	The need to understand metadata standards	Brambilla (2015), Heidorn (2011), Kahn et al. (2014), Newton et al. (2011), Ng'eno (2018), Ohaji (2016), Ray (2012), and Zvyagintseva (2015)
Organisational knowledge	The need to understand policies that govern various aspects of the organisation	Heidorn (2011), Newton et al. (2011), Ng'eno (2018), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Customer relationship	Understanding customer training	Heidorn (2011), Newton et al. (2011), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Interpersonal skills	Interpersonal and communication skills	Heidorn (2011), Newton et al. (2011), Ohaji (2016), Ray (2012), and Zvyagintseva (2015).
Data management	Citing, transforming, editing, describing, and sharing data	Matlatse (2016), Ng'eno (2018), and Zvyagintseva (2015).

RDM training can be achieved in two popular ways according to the literature. First, structured training is the best and quickest way to develop the necessary skills (Brown, et al., 2015). The DCC has already provided practical help of this kind to numerous institutions and individuals (Digital Curation Centre, 2017; Rusbridge et al., 2005) including South African based librarians (Kahn et al., 2014). Second, library schools are being challenged to develop RDM specific courses that could equip their graduates with RDM skill sets and knowledge which would enable them work in research intensive environments.

Ohaji (2016) explored the dimensions of the data librarian role in four universities and five Crown Research Institutes in New Zealand. By collecting data from library and information professionals, repository managers, researchers and information technology service managers, the study found that librarians acquired RDM skills and competencies provided by their organisations in various ways. They included sponsoring them to attend formal learning (university or in-house training); exchange course arrangements with institutions already involved in RDM; attending professional meetings (conferences and workshops); and learning on the job through opportunities (pilot data management, RDM projects and conferences). Partnership training seem to be working perfectly in the popularisation and sharing of RDM skills and knowledge amongst librarians. As mentioned in Chapter One, Kahn et al. (2014) report that the Library and Information Association of South Africa (LIASA) had organised workshops in RDM in partnership with the UK's Digital Curation Centre (DCC) where experts from the DCC shared their RDM knowledge and skills with some LIASA members.

2.6.3. RDM services offered by librarians

As stated already, librarians are increasingly registering their presence in various research disciplines by their active involvement in data curation. These librarians offer various digital curation and preservation services across research disciplines (Walters & Skinner, 2011). Some of these services are discussed in the sections that follow.

Advisory services

According to Brown, et al. (2015), various librarians and research offices may offer various advisory RDM services to researchers. These services may include helping researchers prepare a data management plan for their research grant application, reinforcing the key points of the RDM policy and explaining how the university can help with looking after researchers' data sets.

Data archiving and preservation

There is always demand to store past research data that is no longer actively being used. Some institutions in the UK have developed centralized data centres but have little or no suitable archival storage infrastructure (Brown, et al., 2015). It is further reposted that some of the research institutions are in a dilemma whether to outsource storage or buy in solutions and manage it locally. The hallmark of digital curation is long term storage of research data and there is therefore a need for librarians to fully understand the long-term storage requirements.

While some data sets will be used regularly and should be available very quickly when requested, others will be occasionally requested (Brown, et al., 2015).

Description of data through metadata

Libraries have historically used MARC (Library of Congress, 2013) for encoding metadata with Machine Readable Cataloguinge Xtensible Markup Language MARCXML an updated method of transmitting MARC metadata. Metadata is data about data as it provides more information about the creator of the data and when. According to Musgrave (2003, p. 2), metadata refers to anything that one needs to know to make proper and correct use of the real data, in terms of capturing, reading, processing, interpreting, analysing and presenting the information. It is undeniable that metadata is essential for interoperability, discoverability, provision of information management and also for making data more usable (Brown, et al., 2015; Shakeri, 2013). "Creating data profiles/ data descriptions, which are basically detailed data about scientific data, is a very important part of the data management process" (Shakeri, 2013, p. 10).

Yoon and Schultz (2017) examined RDM services in the USA through a content analysis of 185 library websites. The study identified various services that libraries offered through their websites with data deposit being the most offered services followed by data management planning. Since most websites offered data deposit services, the authors concluded that libraries emphasised promoting and encouraging researchers to participate in RDM activities. What is not however answered in this study is the extent to which researchers utilised these RDM tools on websites to participate in RDM activities. The study noted that, despite most libraries failing to provide basic information for researchers such as the meaning of RDM, about half of the libraries provided at least some information about several core areas of data management including metadata, data preservation and storage, and data publications.

It is reported in the literature that researchers are not very competent in RDM activities and librarians have the responsibility to help them or to curate data for them. For example, Koopman (2015, p. 101) found that although the term metadata is not popular or well known amongst researchers at UCT, some performed metadata functions unknowingly and commonly used the term data description. Nonetheless, some researchers never assigned any metadata to their data. This is where librarians who are naturally information professionals need to register their value by helping such researchers in their curatorial activities.

Librarians with adequate formal and informal RDM skills can impart these skills to researchers. Chen and Wu (2017, p. 352) argue that librarians can provide special training related to research data management focusing on data management and sharing policies; data management plans, data discovery, retrieval and access; format, size; repository requirements; and related tools such as retrieval, recording and processing, preservation and backup for data management and sharing. Latham (2017, p 263) argues that given the libraries' prominence in information literacy and data curation, it is expected that they are in a better position to instruct researchers in best practices for managing their data, creating metadata, and building digital repositories.

2.7. Researchers' data management practices: The voice from the literature

Researchers who are also called original investigators or data producers (Yoon, 2015) have an important role to play in the process of data curation. To highlight their importance in the research data life cycle, the Education-Expertise-Curation (E-E-C) Framework (Bryant, Brian & Malpas, 2017), the Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems, 2002), the CCMF (Lyon et al., 2011, p. 21) and the Digital Curation Centre Lifecycle Model (Higgins, 2008) all underscore the role of researchers in the data curation process. This section discusses various issues related to data management practices by researchers who are the originators of research data.

2.7.1. Data creation and storage practices amongst researchers

One key aspect of research data is to understand where it comes from (Higgins, 2011; Research Information Network, 2008). Scott (2014, p. 121) provides the key ways research is created and collected including scientific experiments, models or simulations, observations and derived data. Walters and Skinner (2011) argue that scholars from various domains are increasingly producing digital information of intellectual value that includes new forms of scholarship, scientific data, notes, electronic records, arts and new media, multimedia learning objects, user-generated web content, and the products of mass digitization efforts. A more comprehensive list of the various categories of electronic data types produced by researchers across disciplines is provided by Scott (2014, p. 121) as follows:

- Textual, e.g. flat text files, Microsoft Word, PDF, RTF)
- Numerical, e.g. Excel, CSV)
- Multimedia, e.g. image (JPEG, TIFF, DICOM), movie (MPEG, AVI)

- Audio, e.g. MP3, WAV, OGG)
- Structured, e.g. multi-purpose Extensible Markup Language (XML), relational (MySQL database)
- Software code, e.g. Java, C
- Software specific, e.g. mesh, geometry, 3D CAD, statistical model
- Discipline specific, e.g. Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) in chemistry
- Instrument specific, e.g. Olympus Confocal Microscope Data Format, Carl Zeiss; and
- Digital Microscopic Image Format (DMIF).

Since the term digital curation was coined, a number of studies have been conducted to examine the role and skills of lecturers or researchers in RDM activities. One such a study was conducted in the USA by Schumacher and VandeCreek (2015) who investigated the status of digital data and preservation practices in five universities. Despite being a developed country with a well-established research infrastructure, it can be concluded from this study that most professors do not have ideal knowledge and skills for managing their digital research data. For example, the study found that most professors followed data management practices that led to the "loss of digital materials that they considered to be important to their professional activities" (Schumacher & VandeCreek, 2015, p. 103). The study showed further that professors used office computers, external hard drives, and flash drives and cloud accounts to store their digital data (Schumacher & VandeCreek, 2015). A related study in the same country (USA) by Shakeri, 2013), showed that research data storage by researchers was extended to cloud based applications where researchers deposited their data for the purpose of storing and sharing with collaborators and students. The study noted further that data was stored in personal computers, hard drives, flash disks and hard discs (Shakeri, 2013, p. 51). Similarly, a study by Housewright, Schonfeld and Wulfson (2013) at the University of North Carolina found that most researchers saved their data in repositories and external hard drives.

A study in China by Chen and Wu (2017) focused on investigating the research data management practices amongst 119 chemistry researchers and postgraduate students at the Chinese Academy of Science. The study, using a questionnaire as a data collection tool, revealed that the most data types produced include experimental data (79.83%) followed by observation data (36.13%). Most data was in capacities of gigabyte level (57.15%) followed by megabytes level (29.41%). In terms of data storage practices, the study found that 81.51%

of researchers used personal computers for data preservation, 74.79% used paper laboratory notebooks and 52.10% of researchers used flash disks or hard disks. Although the study shows that most researchers used personal computers, it can be learnt from this study that laboratory notebooks are still valued as popular ways of storing data implying non-digital research data is still popular amongst Chinese researchers. Most researchers, according to Chen and Wu (2017), attained the basic RDM skills in various ways including special lectures, WeChat, online courses, phone/email, workshops and library blogs. Other researchers have also indicated that the capacities of data stored may range from terabytes, to petabytes, and eventually, exabytes (Walters & Skinner, 2011, p. 63).

Availability of various web based storage facilities have made data storage and sharing much easier. In his study, Shakeri (2013) found that most researchers primarily used cloud based password protected systems to store and share their research data. The common applications, according to Shakeri (2013, p. 41), included email, Dropbox, Evernote, and Google Drive and these applications played a key role in research collaboration.

2.7.2. Reasons for research data sharing, preservation and re-use

The sharing of data is increasingly becoming popular in the research domain mainly because of enforcement by research funders, journal publishers, institutional requirements, data deluge and open access campaigners (Charbonneau, 2013; Chen & Wu, 2017, p. 346; Davenport & Patil, 2012; Guedon, 2015; Matlatse, 2016; Mundel, 2014; Fecher et al., 2015, p. 1). This section discusses various reasons research data should be made available to the public for free.

Teaching resource

Data sharing contributes to science education particularly in training undergraduate and graduate students (Whitlock, 2011). Increasingly, it is becoming difficult for some researchers particularly students to access data for their learning and research purposes. Woolfrey (2009) argues that it is important to make data open so that some groups of researchers such as students, who may struggle to obtain important survey findings through informal channels can re-use such data. Indeed, it may sometimes not be easy for students to be granted permission to conduct research in some organisations that do not understand the value of research hence the only available option is for these students to re-use existing data in their research projects.

Good governance through formulation of evidence-based policies

It is learnt from the literature that sharing empirically generated data can spur sound policy formulation for driving national wide development. Empirical social research can provide the raw material for evidence-based policy (Woolfrey, 2009). The understanding therefore, is that, while research results find their way into public policy through various means, re-examination or re-analysis of initial data is vital as it provides for appropriate policy decisions based on accurate and verified data and research (Woolfrey, 2009).

Cost of research data

Research data produced in academic research teams has become expensive and difficult to reproduce (Davenport & Patil, 2012). Bond-Lamberty (2018), Dai et al. (2018), Kaye et al. (2018), and Shakeri (2013) independently reason that data sharing, re-use and collaboration can minimise the cost and redundancy of research data production. The implication is that when researchers working on similar problems collaborate and share data with one another, redundancies in data collection can be reduced (Shakeri, 2013). Curating data appears to be expensive but other researchers have argued that the benefits supersede the costs incurred. For instance, Rusbridge et al. (2005) argue that the view that data curation is expensive must be well balanced with the economic and social costs of losing digital assets considering that it may be impossible to recreate when they are lost. Moreover, "when data is available, (re-) collection of data is minimised; thus, use of resources is optimised" (Tenopir et al., 2011, p. 1). It is therefore, not surprising that Piwowar (2011) concluded that data re-use is financially sensible.

Foundation for new research

It is generally accepted that knowledge cannot be created in vacuum. Rather, it is created based on prior knowledge (Pisani & AbouZahr, 2010; Watson, 2015). Rusbridge et al. (2005) and Fry, Lockers, Oppenheim, Houghton, and Rasmussen (2008) advocated for research data to be well managed considering that long-term access to the data is crucial in enabling the verification of scientific discovery and in providing a data platform for future research. Other researchers have argued that in addition to acting as a verification tool (Bond-Lamberty, 2018; Dai et al., 2018; Kaye et al., 2018; Shakeri, 2013), data sharing helps in "extending research from prior results" (Tenopir et al., 2011, p. 1). Already, some prior studies have shown that failure to share research data hinders new innovation and discoveries. For instance, in their study which included 1,329 researchers, Tenopir et al. (2011) reported that 60% of the

participants in various disciplines acknowledged that lack of access to data generated by other researchers was a key impediment to innovation and progress in science. One more advantage of data sharing according to Doorn et al. (2013) and Tenopir et al. (2011) is that it adds credibility to the research results in the sense that data availability safeguards against misconduct related to data fabrication and falsification as the available data can be re-analysed to ascertain the validity of the findings.

Rowhani-Farid and Barnett (2016) examined open data practices of 160 researchers who had published in the British Medical Journal between 2009 and 2015. Open data or data sharing is defined by the Royal Society (2012, p.14) as data which is "available, intelligible, assessable and useable". Open data entails making all raw data fully open and available, creating transparency and ensuring reproducibility, and driving further discovery by allowing new knowledge to be generated in the context of earlier discoveries (Pisani & AbouZahr, 2010; Watson, 2015). Findings by Rowhani-Farid and Barnett (2016) showed that only three articles were published alongside their data; 50 authors indicated that they had stored their data elsewhere and were willing to share it with other researchers on request. Generally, the study concluded that the rate of data sharing amongst researchers was very low because of two key reasons: the British Medical Journal has a weak data sharing policy because it leaves room for some researchers not to share their data and secondly, there are no rewards to encourage researches to share their data. For those who shared their data, the main mode of data sharing, according to the study, was through emails. These findings are reinforced by observations made by Acord and Harley (2012) that, unlike sharing published texts through journals, publishing data comes with no tangible rewards.

In a master's study, Zvyagintseva (2015) investigated data management practices in the digital humanities in Canada by analysing the websites of 28 data curation projects taking place in the digital humanities. The study results informed the formulation of an evaluation framework for conducting RDM which spelled out guidelines for establishing metadata standards, data accessibility, and connecting the goals and mission of research projects and RDM practices. However, the study did not investigate the extent of user engagement in the RDM of these projects. The current study bridges this gap by investigating lecturers' involvement in RDM at two public universities in Malawi. In addition, unlike Zvyagintseva's (2015) study which collected data by analysing websites, the current study will use a mixed methods approach (qualitative and quantitative methods) to holistically investigate RDM practices.

In her master's thesis, Koopman (2015) investigated data archiving practices of researchers in the Department of Biological Science at UCT in South Africa. By collecting data through structured interviews, online questionnaires and a literature review, the study found that researchers are increasingly becoming receptive to RDM because 58% of biological sciences researchers archived some of their research in repositories. By learning from previous experiences of data losses, most researchers stored and backed up their data through various ways with the most the popular ones being hard-drives, computers or laptops and the least popular method used was cloud applications (Google Drive and Dropbox) (Koopman, 2015). However, Koopman (2015) fails to show whether UCT offers data storage servers for its researchers or not; the information could have helped to understand better the data storage behaviours of the researchers at UCT.

Most researchers are motivated to use secondary data if it is well managed. Well managed data is characterised as data which contains comprehensive information describing it (metadata), proper documentation and absence of errors (Costello, 2009; Enke et al., 2012; Scot, 2014; Tenopir et al., 2011; Yoon, 2015, p. 173; Yoon & Schultz, 2017). Data re-users will also trust data which is produced by original investigators who have knowledge and experience in data management (Yoon, 2015, p. 144). Woolfrey (2009) argues that metadata should be standardised so that if initial investigators do share their data, secondary analysts need to be able to understand it enough to reuse it. According to Musgrave (2003, p. 8), good metadata should be comprehensive enough to allow secondary users to understand the restrictions which may apply with regard to analysing the data. This is fundamental because Woolfrey (2009) warns that some primary researchers will resort to withholding their complex data sets for fear of inappropriate use by analysts with limited statistical skills. Therefore, the literature suggests that detailed metadata is needed to prevent the misinterpretation of data. A doctoral study was carried out by Yoon (2015) in the USA focussing on a unique topic of the element of trust on data re-users by conducting semi-structured interviews with professors, research scientists and PhD students from the fields of public health and social work. Yoon (2015) discovered that participants were motivated to use existing data because of the cost-effectiveness and potential of secondary data. Before Yoon's (2015) study, and despite the consensus in the professional literature about the importance of data re-use, the concept was poorly researched. Most researchers were motivated to use data if they believed it was from trusted sources implying trust is paramount in data re-use (Yoon, 2015).

In her PhD research study, Woolfrey (2009) chronicled the curation of social survey data in African countries. Social survey data in this context is defined as both the statistical information which is the final product of censuses or sample surveys, and the documentation of the data to facilitate its re-use. The study recommended the development and enactment of data curation policies that can address issues of funding, countering bureaucratic constraints, capacity building, support for professional associations, the establishment of a regulatory infrastructure and development of data infrastructure. The study provided a good general perspective about the trends and status of RDM in Africa. However, since the study relied on websites to solicit data, the current research builds on it by exploring the readiness of RDM in Malawian universities through direct interaction with the researchers and librarians. Also, considering that Woolfrey's (2009) study focused on social survey data, the current study focusses on data in two academic institutions.

Shakeri's (2013) PhD research focussed on understanding data curation activities of researchers at Kent State University' Liquid Crystal Institute which is a small science research unit in the USA. The institute receives limited funding hence the scale of research data production and RDM policies are substantially limited (Shakeri, 2013). According to the study, researchers were mostly interested in sharing their results derived from the data and not the data itself. The study noted that researchers were most willing to share their data with collaborators and students but were less willing to share it with the public. The study provides a comprehensive review of data curation in terms of disciplinary requirements for data management, data characteristics, researcher data management needs and researchers' data management practices for small scale research institutes. However, the study did not extend its scope to the role of librarians in RDM activities. In the current study, in addition to reviewing researchers' involvement in RDM activities, it investigates the role of librarians in RDM as natural partners of researchers.

Since debates about the value of RDM continue to dominate the scholarly literature, researchers have paid particular interest in the concept of data sharing. Despite the advantages of data sharing as orchestrated by some sections of the academic community, many researchers are still reluctant to share their data (Woolfrey, 2009). A reason why researchers pay little attention to data management is provided by Shakeri (2013, p. 71) who argues that researchers are more vigorously involved in day-to-day scientific research and pay little attention to the advantages they may accrue "through sharing, reuse of, and long-term accessibility to their data".

2.8. Data infrastructure: What is its role in data curation?

A good infrastructure is necessary for any RDM project to take off and succeed. Shakeri (2013) argues that data curation cannot be implemented without the establishment of the required infrastructure. Infrastructure in this context refers to integrated systems covering hardware, software and human resources and these are collectively called cyberinfrastructure (Atkins, 2003). In more specific terms, Atkins (2003, p. 5) defines cyberinfrastructure as "software programs, services, instruments, data, information, knowledge, and social practices applicable to specific projects, disciplines, and communities of practice". Worth noting is that infrastructure should not only be perceived to mean or be limited to high performance computing or information technology nor should emphasis only be placed on creating capabilities for data sharing or re-use across the research communities. Rather, Shakeri (2013) proposes additional action goals that fall under cyberinfrastructure and they include acquiring new applications and standards that promote interoperability and that can be incorporated across institutions and disciplines. Well planned and executed cyberinfrastructure is necessary as it will not only ensure accessibility and availability of research data to future generations, but will also make data sharing and research collaboration across researchers from multiple disciplines and at a distance a reality. Infrastructure for RDM purposes comes in different forms**.** A study involving 432 laboratory directors conducted in France by Schöpfel, Ferrant, André, and Fabre (2018) confirm the importance of data infrastructure in RDM activities. The availability of a French National Open Access Repository influenced 70 % of laboratory directors to share data by depositing in this particular repository (Schöpfel, et al., 2018). Researchers need long-term storage but also a short-term version that enables the sharing of active data between research collaborators (Brown, et al., 2015). According to Brown, et al. (2015), a successful technical infrastructure will have three key components: a system for collection, managing and exposing appropriate metadata; a data archive; and a long-term file storage.

2.8.1. Software

Software and data are the most important elements of cyberinfrastructure (Shakeri, 2013). These two elements are interdependent because to store and access data, software is needed. A report by the National Science Foundation of USA (2012) in the USA discusses several software issues in relation to RDM. The report envisions the creation of new types of organisations that integrate library and archival sciences, cyberinfrastructure, computer and

information sciences, and domain science expertise. The report places emphasis on the significance of software development because "it binds together the hardware, networks, data, and users" (National Science Foundation, 2012, p. 4). The report concludes that development of software infrastructure is paramount in supporting data capture and a shared and collaborative system. More importantly, the report argues that good software should have capabilities of supporting current and future expected and unexpected needs. In the UK for example, the Joint Information Systems Committee (JISC) ran a Research at Risk RDM programme that ended on 31 Jul 2016. The project focused on finding and developing solutions for RDM within universities through the provision of a robust and sustainable RDM infrastructure and services to enrich UK research (Brown, et al., 2015). The project helped universities in the UK create, manage and share research data effectively in order to underpin world leading and excellent research (JISC, 2016).

2.8.2. Storage facilities/ data repositories

Data storage is one of the most critical stages of data management in a sense that other data management practices such as documentation, preservation and dissemination are highly dependent on how data is stored (Shakeri, 2013, p. 33). Institutions need to provide storage capacity. In the USA, a study by Schumacher and VandeCreek (2015) categorised data storage practices in three key ways including university-furnished networked storage; free-standing devices, non-networked devices and accounts; and institutional storage without remote network access available. Despite the study reporting about data loss in each of the data storage modes, free standing devices (optical discs and external devices) were the most unreliable and risky. For example, of 31 professors who had experienced data loss, 23 (74%) relied on free-standing devices (Schumacher & VandeCreek, 2015, p. 103). Thus, the study concludes that the use of an institution's network is the best choice and most secure mode of data storage and preservation (Schumacher & VandeCreek, 2015, p. 104).

However, the literature shows that only a few research institutions already have high capacity data storage facilities and those that do not have are responding by buying new storage capacity. The storage capacity required by each institution varies markedly. Some may need up 20 Tb of storage for research data sets for every research project through to the more modest such as 0.5Tb and 1Tb per researcher (Brown, et al., 2015). Brown, et al. (2015) and Walters and Skinner (2011, p. 31) suggested the need for a national data management planning registry to help the higher education community to plan capacity and analyse their progress. In future,

quality, data access, sharing and re-use will depend heavily on how data is stored during the research process. Atkins (2003) emphasises that data management is dependent on the creation of necessary infrastructure and that developing and storing data in an institutional repository assists researchers in better organizing, describing, preserving and providing access to their data.

A study by Ng'eno (2018) which investigated RDM in five Kenyan agricultural research institutes using the CCMF (Lyon, et al., 2011) and the Data Curation Centre's (DCC) Curation Lifecycle Model (Higgins, 2008) revealed that generally, RDM was poorly managed. Ng'eno (2018) noted that there was lack of infrastructure hence data management was not centralised within the institutes as well as at national level. While the study by Ng'eno (2018) investigated the frequency with which researchers made back-ups of their research data, the study did not interrogate the strategies they used. The current study sought to complement this by investigating data back-up strategies employed by researchers in Malawian public universities.

According to Walters and Skinner (2011, p. 31) there are two models for long term digital preservation: those that are implemented centrally and those that are implemented as community driven initiatives. In centrally implemented repositories, the local central library manages a standalone repository whereas in community driven repositories, research universities partner together to set up a preservation infrastructure that caters for the preservation needs of member universities. While the latter cuts across the needs of geographically distributed researchers, the former may be restricted to the users it serves. Another model is where a third party is outsourced or contracted to manage the digital curation activities. Using a third party entity to manage an institution's content brings some challenges such as high costs, limited control over the institution's own data and its general management activities (Walters and Skinner, 2011, p. 31).

2.8.3. Metadata standards and interoperability

Interoperability is defined by Abbott (2009) as the exchange and use of information in an efficient and uniform manner across multiple organisations, systems and platforms. As indicated earlier, metadata plays a key role in data access, sharing and reuse. With regards to interoperability of data sets, it is widely reported in the literature that the absence of metadata standards and formatting standards is a key impediment to sharing and re-using data (Nelson, 2009; Parr, 2007; Sansone & Rocca-Serra, 2012; Teeters, Harris, Millman, Olshausen, & Sommer, 2008; Woolfrey, 2009). The uniqueness of data makes interoperability a challenge as

argued by Callaghan (2013) and Shakeri (2013, p. 74). Callaghan (2013) argue that the nature of data can be unique to a discipline meaning data curation requirements can extensively vary from one another. In that regard, Shakeri (2013, p. 74) highlights that variations in the characteristics of data created in different fields "will necessitate the tailoring of data management practices and plans to fit scientists' needs". Likewise, owing to variations in in data characteristics among scientists, standards and practices developed for large scale research disciplines may not be indiscriminately applied to small research institutes. For example, in systems biology, the Systems Biology Mark-up Language (Hucka et al., 2003; Paton, 2008), is a commonly and widely-used example by most biology scientists. Institutions may use different approaches to metadata collection and management. Some institutions may use existing institutional repositories, others may use their current research information systems while yet others with adequate internal development resources may build their own systems.

Thus, despite having a universally agreed definition, metadata is commonly standardised in each field. This is the case because, according to Shakeri (2013, p. 10), metadata can be created from the perspectives of data producers using their own language. Metadata is necessary for data reusability because they communicate to researchers the validity, value and relevance of the data. This is because metadata provides comprehensive information "about the original data source (e.g. organism, laboratory sample), procedures of how data was generated (e.g. experimental setup, environmental conditions), and further information about unique data attribution" (Wittig, Rey, Weidemann, & Müller, 2017, p. 229). In Microsoft Word documents, the metadata details include the author's name, company and creation/modification date of a document (Microsoft, 2006). According to the Adobe Systems Incorporated (2008), Portable Document Format (PDF) files use two types of metadata: a document information dictionary using a key-value approach and a newer metadata stream stored as XML and complying with the Extreme Memory Profiles (XMP) specification. Marinai (2009) conducted a study that focused on extracting metadata from PDF files for ingestion into digital libraries. The study identified XMP as a metadata specification for image data extraction although increasingly, digital cameras often used the exchangeable image file format (Exif), a type of extraction metadata which can show whether an original image has been modified or manipulated.

Depositing data in a repository requires that it is being accompanied by its metadata. Dublin Core (Dublin Core Metadata Initiative, 2004) are commonly used standards but, depending on the discipline, other metadata standards exist (Scott, 2014). For example, libraries have

historically used MARC (Library of Congress, 2013) for encoding metadata, with MARCXML an updated method of transmitting MARC metadata.

In a doctoral research, Scot (2014) highlighted various issues in managing data and proposed a model for RDM that meets users' diverse needs. The thesis was executed by using data from engineering and medicine. The study demonstrated that differences and similarities do exist in data between disciplines. These differences cut across from types of data created, data management practices to metadata used. Regardless of the standards (Dublin Core, MARC or MARCXML) used to assign metadata to data, Scott (2014) notes that complete and accurate metadata can be useful for data discovery particularly in locating and understanding the data. This is because metadata describes categorises and links data and absence of metadata renders the data meaningless and useless. Without describing and contextualising research data, it becomes challenging to retrieve, access, share, or make sense of research data (Shakeri, 2013); hence it is emphasised that metadata should be assigned to data upon its creation for storage.

Other researchers have argued that the agreed minimum standards of RDM-related metadata are necessary to enable adequate discovery and to support research administration and management (Brown, et al., 2015). Both administrative and preservation metadata are key in RDM efforts. It is desirable to focus on achieving the interoperation of different systems by adopting common metadata standards within or across institutions. This means that many institutions will have to offer a high-level, basic service to researchers that does not account for disciplinary differences in metadata collection although a few will expect researchers to drive the data sets' description process management (Brown, et al., 2015). To make data sharing and re-use easier, Harvey (2010, p. 96) mentions that the "active management of data for current and future use relies on effective sharing of data, which in turn, relies on agreement on and adoption of standards". Similarly, Cragin, Palmer, Carlson, and Witt (2010) are optimistic that data sharing will require research communities to adopt uniform or widely applied data standards as well as disciplinary repository services. Adopting standard exchange formats will culminate in flexibility in automatic and machine-readable data exchange, the development of automatic data workflows between databases, data management systems and applications such as simulation tools (Cragin et al., 2010).

2.9. Impediments to RDM processes

There are a number of data curation challenges that research institutions and universities are facing and certainly dealing with as part of the complex data management activities. This section will discuss a selection of these challenges.

2.9.1. Cost of curatorial activities

According to Walters and Skinner (2011, p. 31), costs associated with ensuring data's accessibility for a very long time are many and high. As e-research and production of other digital intellectual outputs increase in the gamut of terabytes, petabytes, and eventually exabytes, Walters and Skinner (2011) question the capability of universities to build and manage a rich, robust and sustainable organisational storage infrastructure for effective data curation activities. Despite a huge interest from donors to invest in RDM with the aim of making data sets freely and readily available to the research community, it is not easy to obtain such funding. One of the key challenges is the nature of donor policies. Brown, et al. (2015) note that interpreting what funders' policies require with regard to RDM is not always an easy matter mainly because policy requirements can vary from funder to funder, between disciplines and their different priorities sometimes add another level of complexity. For example, the National Science Foundation of USA adopted a data sharing policy a long time ago (Cohn, 2012) but Borgman (2012) reports that it has not been consistent in enforcing the requirements including the aspects of funding. This is perhaps the reasons some research organisations advocate for the need to adopt harmonised local and international data-sharing policies (Denny, Silaigwana, Wassenaar, Bull, & Parker, 2015). Funders do not reward researchers who participate in RDM activities with an understanding that researchers need to willingly engage with the RDM process (Brown, et al., 2015). Some researchers may be willing to share their data but they place their focus on preparing and sharing final research findings because, according to Woolfrey (2009), such initial researchers have few resources to fully document their data. In the light of this, Brown, et al. (2015) recommend the following in relation to funding RDM:

- The need for more work with funders to help universities understand funder requirements and to influence future policy development and implementation
- The need to find ways for creating a reward culture in order to encourage researchers to participate in RDM activities
- The need to develop policies that reflect particular RDM requirements

Although librarians are being heralded as natural executors of RDM, their involvement is limited by a host of challenges. Latham (2017, p. 264) observes that the librarians' roles have been compelled to expand to include an extension into the realm of RDM services but do not always result in additional resources, whether financial or human, allocated specifically for such services or institutional support. However, such problems are rarely experienced in developed countries such as the USA and the UK because research institutions in these countries enjoy better financial support towards implementation of RDM activities.

2.9.2. Lack of expertise

In order to effectively carry out digital curation activities, there is a need to assemble varying bodies of expertise from different professionals (Walters & Skinner, 2011, p. 31). Professionals may come from library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise (National Science Foundation, 2012. p. 4). For instance, information professionals need a deeper understanding of various disciplines in order to efficiently assist researchers in curating research data as well as in creating data by designing ontologies, taxonomies, and creating other forms of metadata. A system that captures, manages and exposes the metadata that describes data sets is an essential part of RDM system skills (Brown, et al., 2015). Yet, Brambilla (2015), Soehner, Steeves, and Ward (2010), and Walters and Skinner (2011) point out that there is generally a lack of skills on the part of information professionals to effectively support curation activities. It has been observed in the literature that most librarians fail to offer RDM services beyond technical services (Latham, 2017, p. 264) yet these skills are in high demand for them to competently handle all RDM issues.

Lack of technical skills in RDM by librarians is attributed to there having been "few professional development opportunities which would afford librarians the competencies necessary to branch out successfully into technical RDM services." (Latham, 2017, p. 264). Fortunately, there are some LIS schools that have introduced formal training in RDM and related concepts. For instance, while students at the University of North Carolina at Chapel Hill in USA have an opportunity to study for a digital curation postgraduate certificate, a master's programme in digital curation is currently offered at Luleå University of Technology in Sweden (Higgins, 2011).

Research on RDM in some African countries such as South African and Kenya has revealed gaps in RDM skills and competencies. In Kenya, a study by Ng'eno (2018) reveals that a lack of skills and competencies in RDM has also affected RDM activities in some Kenyan

agricultural research institutes. A similar challenge (lack of RDM skills and competencies) has been reported in South Africa by Chiware and Mathe (2016), Koopman (2015), Matlatse (2016), and van Deventer and Pienaar (2015). In all these studies, recommendation are made on the need to conduct RDM workshops as a way to equip librarians and researchers with RDM skills. UCT's Department of Knowledge & Information Stewardship (previously called the Library and Information Studies Centre) has been the only institution offering a master's degree in digital curation since 2014 (Library and Information Studies Centre, 2017). In South Africa, there is some positive news about the RDM skills amongst librarians and stakeholders because Chiware and Mathe (2016) and van Deventer and Pienaar (2015) report that knowledge and skills to manage curation activities have expanded and that there is visible commitment from very senior decision makers in the country to build the infrastructure required thereby making eResearch and RDM a reality. The only problem, however, is that in South Africa, librarians are of the view that there is little appreciation for the skills of information management in RDM activities hence the perception that ICT skills are more vital (Kahn et al., 2014, p. 302). This means that librarians should be compelled to acquire advanced ICT skills necessary for RDM activities. Drawing wisdom from these studies, the current study explores the level of skills of librarians and researchers. It further explores if RDM workshops and training are conducted in Malawian public universities.

An important component of expertise in this respect is data literacy. Considering that library schools have introduced courses and modules in data curation, they are being called upon to include data literacy programmes (Prado & Marzal, 2013). In that regard, Chawinga and Zinn (2019, p. 117) reason that "Since librarians are well grounded in delivering information literary programmes, they are equally well placed to deliver data literacy programmes". Prado and Marzal (2013) identify core competencies that researchers need to be data literate. They include understanding data; finding and/or obtaining data; reading, interpreting and evaluating data; managing data; and using data.

2.9.3. Institutional policies

Institutional policies and practices have a great influence on encouraging or inhibiting data sharing. Institutional policies in this context are guidelines developed at institutional level that help dictate which data sets are worthy enough to apply organisational resources to curate them (Walters & Skinner, 2011, p.31). This implies that not all the data produced by researchers may benefit from an institution's resources to curate them. Some institutions may give preference

to curating specific types of data. For instance, health institutions such as medical colleges may give priority to curating data resulting from clinical research and may neglect data resulting from information seeking behaviour studies. In addition, Campbell and Bendavid (2003) noted that some government institutions tend to have a perspective that even data resulting from publicly funded research is very private and develop deliberate policies that restrict sharing such data. Similarly, Denny et al. (2015) observe that although research stakeholders in low- and middle-income settings are aware of data sharing issues, there are no clear policies to guide RDM activities; this hinders RDM implementation. The present study makes attempts to understand the status of RDM policies in Malawi.

2.9.4. Rights management issues

Issues to do with rights management are another notable bailiwick which digital curators need to deal with (Walters & Skinner (2011, p. 31). Beaudoin (2011) reports that librarians' efforts to preserve lecturers' research are punctuated by intellectual property issues. There are two types of rights management in RDM that must be addressed and they include intellectual property rights and data licensing (Shakeri, 2013). Since data cannot be copyrighted, researchers may be reluctant to share data that they perceive is their intellectual property. On the other hand, data licensing refers to the process through which a license, that is a legal instrument for a rights holder to permit a second party to do things that would otherwise infringe on the rights held, is obtained for data sharing and reuse (Shakeri, 2013). The study by Shakeri (2013) showed that researchers were discouraged from sharing their research data (computer codes and software) because they wanted to protect their intellectual property rights and sharing could only be made possible if journal publishers requested the data as supplementary materials to the research manuscripts or papers. In China, Chen and Wu (2017) report that although researchers met the conditions of submitting data to the academic journals, most of them did not meet the requirements of submitting data to the specific data repositories because they were not sure of the copyright issues. Likewise, in the USA, results realised from 1,329 scientists showed that although journal policies compel researchers to share data, these policies "do not necessarily lead authors to make their data sets readily available to other researchers" (Tenopir et al., 2011, p. 2). The other problem is that these journal policies do not address the issue of metadata, long-term preservation, or access for supplementary materials (Dryad, 2010). Data licensing should therefore be addressed before data can be planned for sharing and re-use.

2.9.5. Obsolescence of technologies

Prior studies have identified quite a number of factors that continually and increasingly put research data and other digital information objects at risk. Software and hardware failure was one of the factors that influenced the Consultative Committee for Space Data Systems to develop the Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems, 2002), a digital information preservation model (see next chapter) which for the past two decades, has dominated the archival preservation environments as a reference model. The developers of the OAIS observed that

> *Technology evolution is causing some hardware and software systems to become obsolete in a matter of a few years, and these changes can put severe pressure on the ability of the related data structures or formats to continue effective representation of the full information desired* (Consultative Committee for Space Data Systems, 2002, p. 2).

Schumacher and VandeCreek (2015) report that digital object loss is among other factors largely attributed to digital materials created in past years often not compatible with today's hardware or software. "Technology moves so quickly that there is a real possibility that data collected today will not be readable or even accessible in the future (Kahn et al., 2014, p. 302). More importantly, the storage media of most digital objects is prone to failure especially through accidental damage thereby making binary constructs that make up digital materials lose their integrity (Pogue, 2009). Shen (2016) investigated how data is being stored, managed, shared, and reused by the Virginia Tech faculty and researchers. The study revealed that potential values of data for future research are lost right after the original work is done. A few previous studies (Beaudoin, 2011; Schumacher & VandeCreek, 2015) have reported that most researchers do not understand the causes of digital object loss. The present study will make attempts to find out from the Malawian researchers what they perceive are the contributing factors to the loss of their digital research data.

2.9.6. Poor data infrastructure

Delivering an appropriate level of infrastructure at a cost that is acceptable to the institution is challenging (Brown, et al., 2015). In the UK, only a small proportion of institutions have storage capacity that may be enough for now but they face technical and organisational challenges when it comes to providing an integrated storage solution because their existing infrastructure is distributed across different faculties or multiple sites (Brown, et al., 2015). The

study by Chen and Wu (2017) in China found that most researchers indicated that unreliable storage infrastructure was the key challenges that they faced and this led to frequent loss of their valuable data. In Europe, a study commissioned by the European Union showed that most participants mentioned lack of sustainable hardware, software, and support of the computer environment as the most important threats to digital preservation (Permanent Access to the Records in Europe, 2009). In Africa, particularly in Kenya, Ng'eno (2018) identified poor data infrastructure as one of the key factors confronting RDM activities. Still in Africa, an international qualitative study, covering Uganda and Tanzania targeting 50 professionals working in public health institutions endowed with data, revealed that poor data infrastructure is the consequence of persistent financial constraints (Anane-Sarpong et al., 2017).

2.9.7. Ethical and legal norms

Ethical norms involve the respect for persons involved in the research process. Brakewood and Poldrack (2013) emphasise that issues of informed consent and confidentiality are relevant in most individual related data. In medical or clinical research and other research involving human beings, authors strive to uphold the principle that respondents have the freedom to decide whether to participate or not and may well need assurance that the data will be used only for the purpose of the study (Harding, et al., 2013; Kowalczyk & Shankar, 2011; Mennes, Biswal, Castellanos & Milham, 2013; Sheather, 2009) thereby making it difficult to share or re-use the data. Some data may be described as sensitive and could cause public uproar if it reaches those who may be aggrieved by such data. It is therefore important to consider if the data being shared may inflict harm to the general public or a section of people (Cooper, 2007). In a study by Savage and Vickers (2009) it was found that scientists had concerns about violating patients' privacy (for medical fields), concerns about future publishing opportunities (data re-use), and some were desperate to retain exclusive rights to data that had taken many years to produce. Likewise, Denny et al. (2015) conducted a qualitative study at five research organisations responsible for collecting, curating, and sharing data with the aim of establishing data-sharing practices in South Africa. The study established that informed consent was a key concern when sharing data relating to clinical research data and medical records. Similarly, a quantitative survey involving 10,881 general adults in Japan revealed that "…they [adults] expected stronger protection mechanisms when their family members' clinical and/or genomic data were shared…" (Takashima et al., 2018, p. 1).

Legal requirements may also stymie data sharing. Most common legal issues regarding data sharing include ownership and rights of use, privacy, contractual consent and copyright (Fecher et al., 2015, p. 15) and these are commonly enforced through national privacy acts. Unlike privacy and ethical issues which apply to individual related data, "issues of ownership and rights of use concern all kinds of data" (Fecher et al., 2015, p. 15). In that regard, legal issues concerning ownership of research data before and after deposition in a database are complex and may effectively affect data sharing (Cahill & Passamano, 2007; Chandramohan et al. 2008; Enke et al., 2012).

2.9.8. Terminological differences (digital curation vs. data curation vs. other terms)

Despite the key terms having a common goal and often used interchangeably, they are commonly disciplinary context dependent i.e. data curation for science; digital curation for social sciences and arts and humanities; and digital preservation for libraries (Walters & Skinner, 2011, p. 16). The problem is complicated further by the terminology preference based on cultural aspects where data curation is a naturally preferred term in the USA whereas digital curation is a European preferred term (Zvyagintseva, 2015). Thus, the challenge in discussing this concept is that there is no standardised language (Yoon, 2015, p. 26). The danger of using different terms is that each discipline or region might unknowingly be pursuing separate solutions to manage its digital content yet a "silo-based approach is neither cost effective nor as sustainable as a more unified, campus-wide, and even multi-institutional approach" (Walters & Skinner, 2011, p. 16). Yet, all data curation activities are being supported by the same basic software applications such as databases and mapping tools and features such as metadata, data normalisation, migration of outmoded formats, and stable and sustainable access (Walters & Skinner, 2011, p. 16). To this end, Yoon (2015, p. 29) concludes that "in practice, different domains might curate data differently, but every domain should preserve data using common practices, at least at higher levels".

2.9.9. Lack of incentives for researchers

Building tools for data management does not guarantee its use (Scott, 2014). Compliance alone will not result in researchers embracing RDM willingly partly because the benefits of RDM and long-term storage are hard to sell to researchers and existing incentives are insignificant. The key driving factors are funder and publisher requirements or institutional mandates (Brown, et al., 2015). Funders should commit to covering all costs related to RDM; there is a need for general advocacy for the benefits of RDM (e.g. open data and data mining) among the

wider research community; there is a need to find ways to record downloading information and other statistics of shared or archived data so as to encourage researchers to engage with RDM (Brown, et al., 2015). Data curation advocacy could include the introduction of incentives for the free sharing of research data (Woolfrey, 2009, p. 115). More importantly, effective promotion of the standard method for citing data sets and encouraging researchers to re-use and cite other people's data sets could both be important (Brown, et al., 2015). Data citation is already motivating researchers in some Arab Universities (Egypt, Jordan, and Saudi Arabia) to share data as revealed by an online study involving 337 respondents by Elsayed and Saleh (2018). The study showed that 64.4% of researchers shared their data hoping to be cited thereby remaining visible in their academic disciplines. Findings reported by Elsayed and Saleh (2018) are therefore, not puzzling considering that, in the words of Patterton, Bothma, and van Deventer (2018, p. 22), "Researchers are more easily persuaded to add their data to a repository when they know the data would be cited – just as their articles are". More importantly, Aguillo, Bar-Ilan, Levene and Ortega (2010) and Soh (2015) report that the QS World University Ranking and the Times Higher Education World University Rankings which are the highly valued university ranking systems, take into account publication and citation counts when ranking universities.

2.10. Building data curation centres: Two case studies

As of 2012, Cox and Pinfield (2016) found that UK academic libraries offered limited RDM services to researchers. However, the study concluded that data management, external archives, citation, copyright and licensing were beginning to emerge as priorities. To underscore the importance of digital curation, in addition to early curation centres such as the Digital Curation Centre which was launched in 2004 (Higgins, 2011), several notable others have been established recently. They include the Greek Digital Curation Unit (Athena Research Centre), the University of California Curation Centre, the Digital Research and Curation Centre (Johns Hopkins University's Sheridan Libraries), the Digital Curation Institute (University of Toronto - Canada), the Purdue University Library's Distributed Data Curation Centre, eResearch Centre at University of Cape Town and German Network for Bioinformatics Infrastructure (Chiware & Mathe, 2016; Higgins, 2011; Walters & Skinner, 2011; Wittig et al., 2017, p. 229). An analysis of the literature suggests two complementary fundamental remits of these digital curation centres: to assume complete and direct control over the management, curation, and preservation of the information resources underpinning their scholarly activities; and to

develop automated tools and effective workflow that ensure long-term access and preservation to both digitised and born-digital materials.

In the section that follows, the researcher will review two digital curation centres which include the de.NBI Systems Biology Service Centre (Germany) and the DCC in the UK. The de.NBI Systems Biology Service Centre is selected because it is an example of a curation centre that offers its RDM services to a unique group of researchers (biology researchers). The UK DCC is selected because of its popularity across the globe in offering RDM expertise despite not being a repository itself.

2.10.1. The de.NBI Systems Biology Service Centre

The de.NBI-SysBio is a systems biology data management service centre within the German Network for Bioinformatics Infrastructure (Wittig et al., 2017, p. 229). Services offered by the centre include application users, developers and expert data analysis (German Network for Bioinformatics Infrastructure, 2017). Services to users include access to manually curated and enriched curated data where users can search, store and exchange their data, and models and operating procedures. The centre also trains users in encoding and publishing their results in reproducible manner. Services to data developers include offering guidance on setting up their own RDM projects and training them in how to use and extend the SEEK software which the centre uses to manage its data. The centre also offers assistance in the implementation of systems biology standards for developers' own simulation and analysis tools. The centre provides expert data analysts with web services for automated data search, modelling and simulation tools and access to reproducible modelling results. The data management practices at the centre are guided by the FAIR principles where the letters in FAIR stand for Findability, Accessibility, Interoperability, and Reusability respectively (Wittig et al., 2017, p. 229). Since the other terms are defined elsewhere in this document, it is important to define Findability for clarity. Findability entails the use of standard identifiers and annotations which point to standard ontologies and databases as well as the use of controlled vocabulary. Findability aims at solving unambiguously access to data by making sure that there is consistency in the use of identifiers and vocabularies across documents and vocabulary. A review of the centre's RDM activities by Wittig et al. (2017) showed that the data is available to the general public and can be accessed and displayed within the web interface as long as the file format is supported (e.g. Excel, Word, PDF) and can be downloaded and saved to local machines such as computers

and smartphones. The key challenge has been to convince researchers to view RDM as a natural extension to publishing research findings in the scientific community.

2.10.2. Digital Curation Centre (UK)

The Digital Curation Centre is an interdisciplinary data curation centre which was launched in 2004 (Higgins, 2011, p. 81) in response to a call for the realisation that digital information is both essential and fragile (Digital Curation Centre, 2017; Rusbridge et al., 2005, p. 1). Its core service is the provision of expert advice and practical help to research organisations wanting to store, manage, protect and share digital research data. Upon its establishment, the centre aimed at accomplishing various core roles including championing an understanding of digital curation issues among scholars and researchers; offering services that promote digital curation; sharing knowledge and skills in digital curation across research disciplines; developing and offering technological support for digital curation; and conducting long-term research on digital curation (Rusbridge et al., 2005, p. 1). From 2011, the DCC has worked with various organisations to develop and implement tailored RDM services and projects (Digital Curation Centre, 2017). It has further offered consultancy services, training programmes, policy development and data management plans. In Africa, the DCC offered RDM training at a workshop organised by the LIASA where library professionals were exposed to and equipped with RDM skills and knowledge (Kahn et al., 2014).

The DCC is not itself a digital repository. Instead, its role is to offer innovative services and guidance that enable data centres and repositories to be more productive in performing data curation activities (Rusbridge et al., 2005). The DCC promotes RDM activities through various means. Firstly, it publicises its services and makes resources available via a web portal or website where RDM stakeholders can access their resources for free. Secondly, the DCC produces an electronic journal called the International Journal of Digital Curation which is dedicated to digital curation and preservation research. This e-journal does not promote the activities of DCC but rather, it is a highly rated platform where scientists present their contribution to the field of digital curation (Rusbridge et al., 2005). Finally, the DCC engages with various partners through its Associates Network Initiative (Higgins, 2011, p. 81). Through this initiative, DCC brings together key members "from the UK data creating and managing organisations, leading data curators overseas, supranational standards agencies, and representatives of sectors of UK industry and commerce involved in digital curation" (Rusbridge et al., 2005, p. 10).

2.11. Summary of chapter

This chapter reviewed research related to the concept of RDM practices. The chapter reviewed the literature from various countries across the globe. Specifically, the literature reviewed was generated from the UK, USA, China, Canada, Germany, Australia, South Africa, and Kenya. From the literature searched and reviewed, it was clear that RDM is an emerging concept and has been profoundly popularised and adopted in European countries. It was therefore not surprising that the bulk of the literature reviewed in this chapter was generated in European countries or by European researchers.

From the reviewed literature, it was established that data curation is an emerging role that has to be embraced by librarians. These roles, according to the literature, are both service based and technology based implying that much as librarians have for centuries served as information service providers to researchers, they need to master more advanced ICT skills to effectively perform data curation activities. It was further established that some LIS schools across the globe have started offering formal training in data curation thereby exposing librarians or information professionals to these new skills. The literature also showed that workshops on data curation are key to equipping librarians with RDM skills and knowledge. It was noted that the UK DCC has been playing a central role in skills development for librarians through workshops and development and provision free data curation resources. In Africa, it was noted that South Africa is the only country with tangible initiatives for RDM.

The reviewed literature showed that RDM allows the reproducibility of study results and the re-use of old data for new research questions. It is therefore plausible to conclude that open sharing and re-use of research data reduces costs through collaboration amongst researchers; leads to new discoveries based on existing or previously generated research data; leads to reduced duplication of data; leads to increased transparency of the research record through verification; and leads to an increased and immediate research impact.

The literature also unveiled that researchers are key in the process of data curation because they are the creators or originators of research data. Although most researchers are in the developed world such as the USA, UK, Canada and Germany, and produce substantial amount of research data, not many are willing to share their data for various reasons such as fear of data misinterpretation. However, despite their resistance to sharing their data, researchers are gradually succumbing to pressure exerted by research funders, journal publishers and the open access movement and are beginning to embrace the concept of open data. In addition to using

personal computers and flash discs (memory sticks), there was little evidence of the use of institutional infrastructure as data storage facilities. Some researchers used cloud based application such as Google Drive, Dropbox and emails to store and share their data. In Africa, the available literature about data curation was mostly generated in South Africa.

Several research works discussed the aspect of cyberinfrastructure necessary for executing data curation activities. All the reviewed research emphasised having a robust infrastructure including software, hardware and metadata standards. Many studies indicated that metadata standards are key to smooth sharing and repurposing of research data across disciplines. However, regardless of a consensus in the literature about the need for uniform metadata standards, it was shown that many research disciplines continue to develop discipline specific standards. The consequence has been that a silo-based approach to metadata is neither cost effective nor as sustainable because instead of adopting available solutions, each discipline or region ends up unknowingly pursuing separate solutions to manage its digital content. Some researchers are more optimistic by stating that considering the fact that variations in metadata standards are costly and limit data sharing and re-use, in the near future there is a high possibility of developing a unified and multi-institutional approach to metadata standards.

The literature showed a number of factors that stymie the popularisation, adoption and implementation of data curation projects. Cost associated with curating data is a mountain to overcome because several resources are required to implement data curation projects. For example, due to high costs, there was generally poor cyberinfrastructure such as inadequate storage facilities as reported by some researchers from China and the UK. Other challenges cited in the literature include lack of RDM expertise on the part of librarians and researchers; ethical and legal norms; lack of incentives for researchers; obsolescence due to rapid changes in technology where digital assets created in the past become incompatible with new hardware; unpalatable RDM institutional policies; and terminological differences where it was noted that various disciplines use different terms to refer to RDM. Finally, two evidence based interventions were reviewed namely, the UK DCC and the de.NBI Systems Biology Service Centre in Germany.

From this comprehensive literature review, it was observed that no literature was found in Malawi about RDM implying that nothing is so far known about the popularisation, adoption and implementation of this emerging concept. However, one of the core functions of universities is to conduct research and this implies that research teams and individual

researchers in public universities are producing research data. With the proliferation of ICT hardware such as laptops, computers, flash disks, mobile phones, digital cameras and the Internet in universities in Malawi, it is plausible to suggest that researchers are utilising these applications to create digital research data. It is for this reason that this study was conducted to explore RDM practices in two public universities in Malawi. In the next chapter (Chapter Three), the researcher will present the conceptual and theoretical frameworks underpinning the study.

CHAPTER THREE

THEORETICAL FRAMEWORKS

3.1. Introduction

Since the concept of data curation was conceived more two decades ago, researchers have endeavoured to develop models that guide the data curation processes in various organisations. In this chapter, the researcher reviews some models that underpin research in the data curation and related fields. Four models will be discussed focusing on various aspects. The models include the Education-Expertise-Curation (E-E-C) Framework (Bryant et al., 2017), the Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems, 2002), the CCMF (Lyon et al., 2011, p. 21) and the DCC Curation Lifecycle Model (Higgins, 2008). The latter two have been adopted to guide this study.

3.2. Why should research be guided by theoretical frameworks?

When undertaking research, it is crucial to identify an appropriate conceptual framework which will guide the study (Blanche & Durrheim, 1999; Lesser, 2000; Noko & Ngulube 2015). Researchers formulate theories to help explain, predict and understand issues being studied as well as to challenge and extend existing knowledge within the limits of critical bounded assumptions (Abend, 2008; Blanche & Durrheim, 1999; Grant & Osanloo, 2014; Noko & Ngulube 2015; Swanson & Chermack, 2013). Some researchers (Corvellec, 2013; Jaccard & Jacoby, 2010; Maxwell, 2012; Ravitch & Riggan, 2016; Trochim & Donnelly, 2006; Sutton & Staw, 1995) have explained the reasons why research should be guided by theoretical frameworks as follows:

- Theoretical frameworks lead and connect researchers to existing knowledge. The adopted theory helps the researcher to justify the choice of research methods.
- Theoretical assumptions guide researchers in addressing the *why* and *how* questions thereby permitting the researchers to intellectually transition from simply describing the phenomenon based on what has been observed to generalising various aspects of that phenomenon.
- Theoretical frameworks help researchers to identify the limits of study generalisations because theories specify the key variables that influence a particular phenomenon.
- Theories provide members in a particular field of study with a common language and a frame of reference for demarcating the boundaries of their profession.

- Theories guide and inform research in such a way that outputs from such research improve professional practice.

Thus, a theoretical framework is a 'blueprint' for the entire dissertation inquiry which serves as a foundation on which the researcher "philosophically, epistemologically, methodologically, and analytically approach[es] the dissertation as a whole" (Grant & Osanloo, 2014, p. 13). Generally, theoretical frameworks provide researchers with clues or guidelines to answer the questions which researchers may simply speculate or fail to offer any conclusive explanation.

3.3. The Education-Expertise-Curation Framework

The Education-Expertise-Curation (E-E-C) Framework was an outcome of an Online Computer Library Centre commissioned Report (Bryant et al., 2017). The report was prepared by Bryant et al. (2017) based on a study whose focus was to examine the infrastructure, services and resources that the higher education in the developed world needed to support RDM practices. Four research intensive universities, the University of Edinburgh (UK), the University of Illinois at Urbana-Champaign (USA), Monash University (Australia) and Wageningen University and Research (the Netherlands), were included in this research project. The E-E-C Framework is mostly concerned with RDM services and it consists of three distinct RDM service categories which include *education services*, *expertise services* and *curation services* (Bryant et al., 2017). These services are presented in Figure 1.

3.3.1. Education services

According to Bryant et al. (2017), under this service category, the aim is to educate researchers or the university scholarly community and other stakeholders about various issues concerned with RDM. Such issues include the importance of managing their research data to ensure that it is available for future use by themselves or other researchers. Another issue of interest under this service category is the need to educate researchers about the relevant obligatory RDM polices imposed by funding organisations, local or international agencies and researchers' own research institutions. This service category of E-E-C aims to equip researchers with basic data management skills such as developing reasonable data management plans and guidelines for creating descriptive metadata that necessitate discovery and re-use of archived data sets. Bryant et al. (2017) concur with many others scholars (Chen & Wu, 2017, p. 346; Enke et al., 2012; Kim et al., 2013, p. 67; Wallis et al., 2013) that compliance with funder requirements is the key incentive for data sharing amongst researchers but identify an additional form of incentive,

reputation. They argue that researchers and their institutions will enjoy enhanced reputations for sharing their data sets for re-use.

3.3.2. Expertise services

The focus of this service category is to provide researchers with customised support and solutions related to RDM services. Such support services may include providing appropriate means through which researchers can channel relevant RDM questions; direct consultation with data or liaison librarians; and offering customised RDM support such as metadata creation and mediated deposit (Bryant et al., 2017). Two distinguishing features of this service category are noted. First, these services are offered through direct interaction between the researchers and data management experts, that is, there is no use of unmediated tools such as LibGuides or self-paced online tutorials. Second, expert services operate in parallel with the research process itself meaning they are offered to researchers at any stage of the research cycle. This is in contrast to the education services (discussed in the previous sections) which can be consumed by researchers independent of any particular research stage (Bryant et al., 2017). The understanding is, therefore, that at this stage, researchers lack RDM skills and knowledge and consequently, they encounter problems in managing their data hence, they seek assistance from data librarians and technologists who boast of vast RDM expertise and knowledge.

3.3.3. Curation services

The focus of this service category is on the infrastructure functionality that is needed to manage data throughout the research cycle. It includes active data management during the research process as well as long-term stewardship of data, that is, the careful management of data for repurposing after the research process has been completed. To achieve long term data stewardship, Bryant et al. (2017) propose key functions that are performed in this service category. These functions include storing data, assigning unique identifiers to data, controlling access to data, metadata creation and long term preservation. These functions can be implemented as short term, medium term or long term. According to Bryant et al. (2017), putting infrastructure in place alone is not enough because policy issues come into play when implementing curation services. For example, retention policies may dictate which data sets will be deposited (appraisal), which data will be removed after a prescribed period of time and which data will be retained indefinitely. These retention polices will also ensure compliance with requirements imposed by funders, national and international agencies. Other important

RDM policy areas under curation services include metadata requirements, access restrictions and privacy assurances for sensitive data (Bryant et al., 2017).

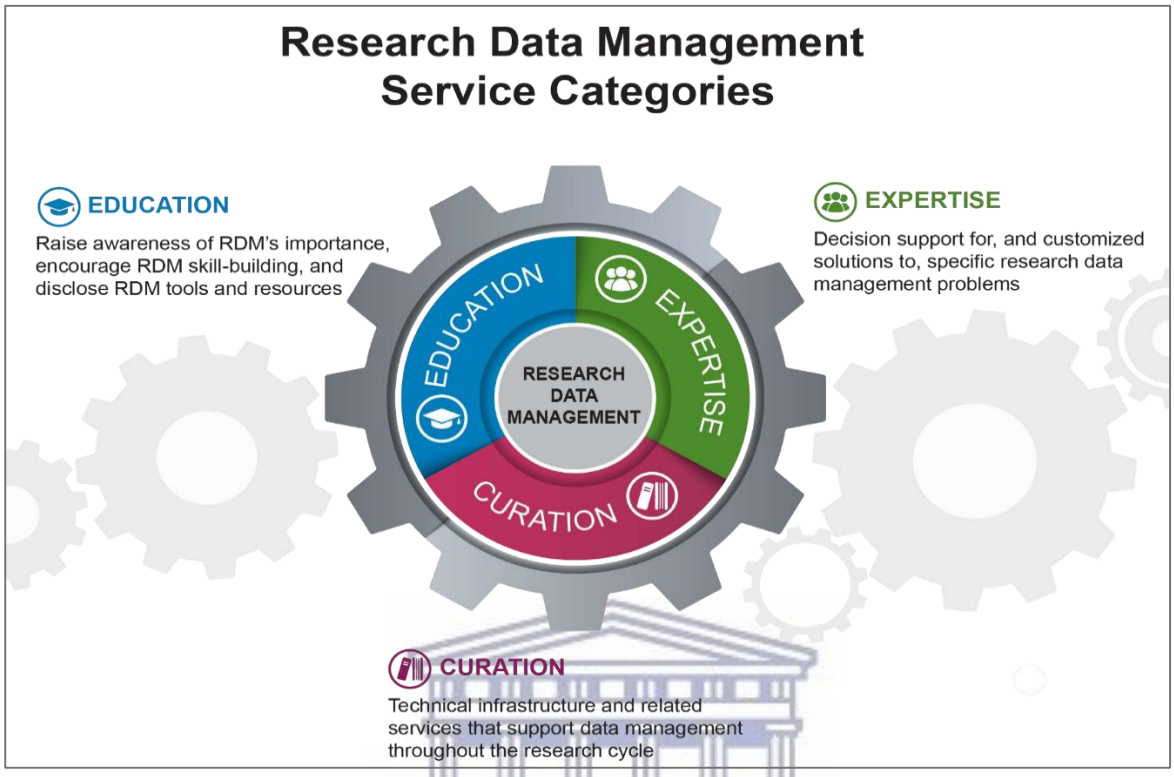


Figure 1. The Education-Expertise-Curation Framework (Bryant et al., 2017)

3.3.4. Strengths, application and shortfalls of the theory

This model is less than two years old and there would be little expectation that it has been adopted, applied and rigorously verified by other researchers. But being new does not mean the model should be neglected by other researchers. However, an analysis of the model reveals some general strengths worth noting. Like other well verified curation models, through the education services, the model attempts to engage the scholarly community on the benefits of curating data. This is important because raising awareness is an important step towards winning the participation of researchers in RDM practices. The implication of this strength is that the model can be used even in research institutions where RDM services are not yet introduced but plans are in the pipeline. Another strength of the model is that it retains key attributes of the commonly used models in data curation. For instance, it emphasises the core curation aspects such as technical infrastructure which is the central key issue covered in comparable but rigorously verified and highly rated models such as the Open Archival Information System (OAIS) and the Digital Curation Centre's Curation Lifecycle Model (Higgins, 2008).

When appraised against other conventional curation models, some shortcomings can be isolated from the model. Although this may not be a good strategy for rating the appropriateness of the model, there are some universal and core theoretical aspects that need to conform to a particular field or domain. Theories offer members in a particular field of study a common language and frame of reference for demarcating the boundaries of their profession (Corvellec, 2013; Jaccard & Jacoby, 2010; Maxwell, 2012; Ravitch & Riggan, 2016; Trochim & Donnelly, 2006; Sutton & Staw, 1995). One key weakness of the model is that it does not provide any systematic stages from research data creation to data re-use. This leaves the reader or potential users wondering what comes first amongst education, expertise and curation services. The developers could have used either connectors (arrows) to demonstrate the flow of activities amongst the three parameters or they could have highlighted interconnectedness and interdependence of the three service categories in their narration. The model regards researchers as novices in RDM but regards librarians and IT personnel as experts who should mostly be charged with the responsibilities of perpetually helping researchers. Although this is acceptable, the worry is that researchers may become over dependent on data librarians and technology experts. A modified model of the same may consider including an aspect on how to equip researchers with basic RDM skills for managing their research data.

3.4. Open Archival Information System (OAIS)

The OAIS is an International Organization for Standardization (ISO) conceptual reference model designed to inform the development of systems for the long-term preservation of digital information (Consultative Committee for Space Data Systems, 2002; Lee, 2015, p. 25). Like other digital preservation models, the key aim of the model is to enhance a broader understanding of the requirements for long term preservation and access to information, particularly digital information. It is important to mention upfront that, the OAIS uses the term information as a hypernym for various types of information that are exchanged and managed in an archival system. These types of information include data and knowledge (Habert & Huc, 2010., p. 426). In more specific terms, according to Consultative Committee for Space Data Systems (2002, p. 1), the OAIS was developed to achieve six aims: to provide the basis for raising awareness of archival concepts and technologies needed for long-term preservation of information; to educate and raise awareness for digital information preservation procedures; to provide common ground for discussing various long term preservation strategies and techniques; to provide the basis for making comparisons about various forms of data models preserved by repositories; to act as a platform on which efforts to improve long-term

preservation practices for non-digital information can be made; and to offer guidelines for identifying and developing OAIS related standards.

The OAIS consists of six distinct but interdependent functional entities. See Figure 2. They include access, administration, archival storage, common services, data management, ingest, and preservation planning (Consultative Committee for Space Data Systems, 2002; Lee, 2015, p. 25). There is an interplay among three distinct features in the OAIS: Submission Information Package (SIP), Archival Information Package (AIP) and Dissemination Information Package (DIP) (Consultative Committee for Space Data Systems, 2002; Lee, 2015, p. 25). According to Consultative Committee for Space Data Systems (2002, p. 10), SIPs are what the OAIS receives or collects from producers; AIPs are what the OAIS manages and preserves and DIPs are derived from one or more AIPs which are received by the consumer in response to a request to the OAIS. This means that information enters the OAIS as SIPs, is processed and managed as AIPs and it exits the OAIS as DIPs.

3.4.1. Ingest

It is the OAIS entity that contains the services and functions that accept SIPs from producers, turns them into AIPs by preparing them for storage and preservation (Consultative Committee for Space Data Systems, 2002). Specific functions of this entity include receiving SIPs; performing quality assurance on SIPs; generating an AIP that complies with the archive's data formatting and documentation standards; extracting Descriptive Information from the AIPs for inclusion in the archive database; and coordinating updates to Archival Storage and Data Management (Consultative Committee for Space Data Systems, 2002).

3.4.2. Archival storage

It is an OAIS's entity that contains the services and functions for the storage, maintenance and retrieval of AIPs. This entity receives AIPs from Ingest; adds them to permanent storage, manages the storage hierarchy, refreshes media holding archives; routinely checks special errors; provides disaster recovery capabilities; and provides access to AIPs (Consultative Committee for Space Data Systems, 2002; Lee, 2015).

3.4.3. Data management

This entity is responsible for populating, maintaining, and accessing a wide range of information. Examples of its functions include populating the information of catalogues and inventories with what has been retrieved from archival storage; processing algorithms that may

be run on retrieved data, consumer access statistics, consumer billing, event based orders, security controls, and OAIS schedules, policies, and procedures (Consultative Committee for Space Data Systems, 2002). The entity also supports access to descriptive information (Lee, 2015).

3.4.4. Administration

This entity supports the services and functions for the overall operation of the archive system. Functions falling within this entity include soliciting, negotiating, and submitting agreements with producers; ensuring that submissions meet archival standards; configuring system hardware and software; providing engineering function necessary for archival operations; and migrating and updating contents of the archive (Consultative Committee for Space Data Systems, 2002). Another key function of this entity is to establish and maintain archive policies and to activate stored requests and providing customer support (Consultative Committee for Space Data Systems, 2002).

3.4.5. Preservation planning

The key function of this entity is to provide services and functions for monitoring and maintaining the environment of the OAIS. It provides recommendations that aim at ensuring that information stored in the OAIS is accessible to the designated user community for as long as it is required despite obsolescence of the original computing environment. This entity evaluates the contents of the archive thereby developing recommendations for updating archives by migrating current archive holdings; developing recommendations for archive standards and policies; developing detailed migration plans, software prototypes and test plans that enable implementation of administration migration goals (Consultative Committee for Space Data Systems, 2002).

3.4.6. Access

This entity connects the user community with the OAIS by providing services and functions that support the existence, description, location and availability of information stored in OAIS (Consultative Committee for Space Data Systems, 2002) implying that users use this entity to request and retrieve information products stored in the OAIS. Specific functions performed by this entity include interacting with consumers to receive requests; applying control to limit access to protected information; coordinating the execution of requests to successful completion; generating responses; and delivering the response to users.

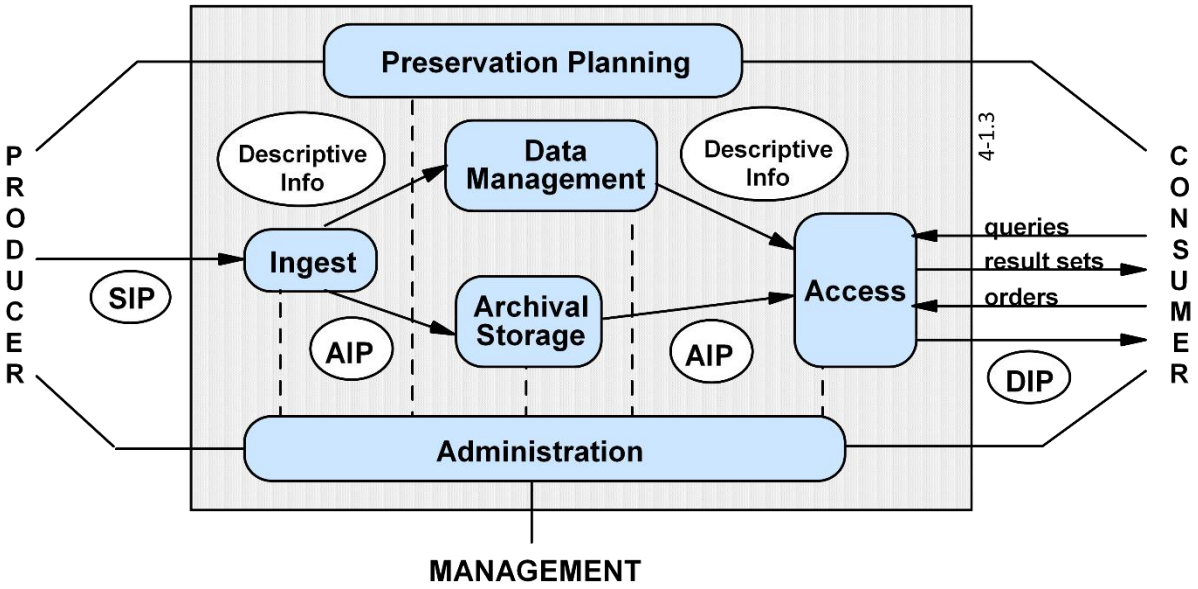


Figure 2. The OAIS Model (Consultative Committee for Space Data Systems, 2002)

3.4.7. Strengths, application and shortfalls of the theory

The AOIS is one of the oldest models conceptualised to address the debates about the long-term preservation of digital objects. There are several strengths in this model and some are worth highlighting. One notable strength of the OAIS model is that it can be used to design and develop software (Egger, 2006) that is necessary for the preservation of digital content. Another well proven strength is that the OAIS is a flexible conceptual reference model that can be customised or transitioned into actual practice that fits different information preservation situations (Vardigan & Whiteman, 2007), meaning it does not "provide [rigid] implementation guidelines" (Dunckley, Reshef, Conway & Giaretta, 2010, p. 81). Rather, it simply aims at identifying and describing core strands that assist curators to effectively plan, coordinate and implement the preservation of digital content for longevity (Lavoie, 2004, p. 77; McDonough, 2012). Another strength is that the model has significantly influenced the development of other data curation models such as the CCMF (Lyon et al., 2011, p. 21) and the DCC Curation Lifecycle Model (Higgins, 2008, p. 136). It is not common to notice features in most data curation models that are a direct replication of the OAIS.

The model is however not short of criticism from the scholarly community. One of the weaknesses is that the model's integration of the management and technical functionality entities makes it difficult for software developers to design and develop an archival management system that can effectively accommodate both, the management and technical

aspects (Egger, 2006). In addition, McDonough (2012, p. 1631) observes that like other models which are mostly theory based, using the model to conduct long-term preservation of some digital information media such as computer games can sometimes be problematic because "theory and practice can sometimes have an uneasy fit".

Considering that the shortcomings of the model are insignificant, the OAIS has been adopted and used by various scholars and professionals in digital preservation projects and research activities. Lee (2010) and Greenstein and Smith (2003) report that the OAIS has been widely used as the basis for research and development in data curation whereby most conference papers, articles and reports by researchers are increasingly being presented within the context of the OAIS framework. More commonly, the OAIS has been widely used to create and maintain digital repositories (Jeng, He, & Chi, 2017, p. 626; Laughton, 2012, p. 308; Lavoie, 2004, p. 70; Lee, 2010; Ray, 2012).

Egger (2006) used the OAIS to design, develop and implement the Austrian Literature Online Digital Repository Software (ADIGRES). Using the OAIS, Jeng et al. (2017) examined the Inter-university Consortium for Political and Social Research (ICPSR) focussing on general operational issues such as responsibilities, ICT practices, and barriers and challenges. It was observed that the "OAIS model is robust and reliable in actual service processes for data curation and data archives" (Jeng, et al., 2017, p. 626).

3.5. The Community Capability Model Framework

The CCMF is a not an old theory which was developed by Lyon, et al. (2011). The CCMF was developed as a self-assessment tool for disciplinary researchers (Lyon et al., 2011; Qin, Crowston & Lyon, 2016). The Framework comprises eight *capability factors* that represent human, technical and environmental issues. Within each of these factors, are a series of *community characteristics* that are relevant for determining the capability or readiness of that community to perform data intensive research (Lyon et al., 2011). The eight capability factors are as follows: collaboration; skills and training; openness; technical infrastructure; common practices; economic and business models; and legal and ethical issues. Figure 3 presents the CCMF.

3.5.1. Collaboration

Collaboration in context entails the working relationship that takes place at different levels of research. Collaboration, according to Lyon et al. (2011), may be categorised as lone

researchers, departmental research groups, cross-research group interaction, discipline organised at a national level; and international collaboration and consortia. Collaboration is vital in research because working relationships that are formed during research have a strong bearing on the types and value of research being carried out (Lyon et al., 2011). In the higher education context, researchers may form alliances with stakeholders such as public and private institutions and industries to collaborate in solving problems (Qin et al., 2016).

3.5.2. Skills and training

According to Lyon et al. (2011), the capability of the community to perform intensive research is strongly influenced by the individual capabilities of its members. Community capability can therefore be enhanced by training its members in the relevant skills. In this context, such skill sets may focus on "tools and technologies (cloud computing, visualisations, statistical analysis, simulations, modelling); data description and identification (metadata, vocabularies, citation); and policy and planning (data management, business models)" (Lyon et al., 2011, p. 25).

3.5.3. Openness

It refers to the open communication of research methods and results which, according to Lyon et al. (2011), leads to scientific progress. The principle of openness applies at different levels of research: communicating the plans for research, communicating ongoing research progress being undertaken, and opening up the published literature to a wider audience (Lyon et al., 2011, p. 25). In order to add value to the whole research process through validation, reproducibility and reusability of research calls are increasingly being made by the research community to open up the data, methodologies employed alongside final results and conclusions (Guedon, 2015; Lyon et al., 2011).

3.5.4. Technical infrastructure

The technical infrastructure supports research tools and services that are used at different stages of the research life cycle (Lyon et al., 2011). Examples include computational tools and algorithms; data capture and processing, storage, curation and preservation; data discovery and access; platforms for integration, collaboration and citizen science and visualisation and representation (Lyon et al., 2011). Computer based storage facilities are necessary for data and research management. For example, The University of Bristol has invested in petabyte-scale research data storage that supports the large scale data sets being produced by researchers (Lyon, et al., 2011).

3.5.5. Common practices

The argument, according to Lyon et al. (2011), is that research communities have produced *de facto* standards in some areas such as data formats, data collection methods, processing workflows, data packaging and transfer protocols, data description, semantics, ontologies and vocabularies, and data identifiers. These common standards need to be shared and understood by researchers within a particular research field.

3.5.6. Economic and business models

Lyon, et al. (2011) are of the view that data-intensive research requires some degree of investment, and it is therefore important to consider how this might be sustainably funded. Investments focus on two aspects: research (such as major investment in longitudinal data surveys in the social sciences) and infrastructure (such as large central investments in network infrastructure).

3.5.7. Legal and ethical issues

In research, there may be ethical obligations that might obstruct sharing of certain datasets and, legal issues may act as barriers to sharing data in the first place and repurposing or re-using these data. In medical research for example, "consent forms signed by patients strictly limit what can be done with the data later on" (Lyon, et al., 2011, p 45).

3.5.8. Academic issues

The argument put forward in the CCMF is that the research activities being carried out by the academic research community should be recognised. Lyon et al. (2011, p. 42) argue that "intensive research is most likely to flourish in communities where data is valued highly: where researchers are rewarded for their data contributions, and high standards are expected of data entering the research record". One of the successful reward models, according to the CCMF, is where all contributions by researchers are recognised and rewarded, through established procedures and measures.

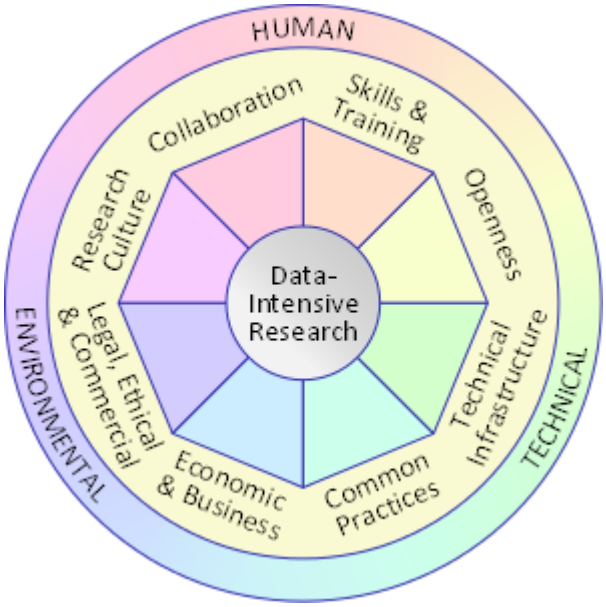


Figure 3. The Community Capability Model Framework (Lyon et al., 2011)

3.5.9. Strengths, applications and shortcomings of the theory

By analysing the model, it is clear that it cross-cuts two highly interdependent research aspects namely, the actual research process and the previously neglected aspect of RDM. Thus, unlike other comparable models, the CCMF provides a balanced description of the factors that enable the creation of research data and its preservation.

Owing to its relevancy, the CCMF has been successfully used by three prominent studies to examine RDM in Europe, USA and Africa. In Africa, Ng'eno (2018) used the model to investigate the research data management practices at some selected agricultural research institutes in Kenya. The model helped to understand various variables in research data management in research institutes in Kenya, namely data openness; skills and training; technical infrastructure; legal and policy issues; and research collaborative partnerships. In the USA, Shen (2016) used the CCMF to analyse data sharing activities amongst researchers at Virginia Tech. The study revealed that open data sharing was not adequately embraced by researchers due to various factors, such as lack of funding, limited time and absence of incentives (Shen, 2016). In Europe, Lyon, Patel, and Takeda (2014) used the CCMF to identify the requirements for RDM in academic libraries by gathering data from eScience researchers who participated in an international workshop from Cambridge (UK), Melbourne (Australia), Stockholm (Sweden), Bristol and York (UK), and Amsterdam (Netherlands). Based on the study findings, the researchers developed summaries and visualisations of data-intensive

capability which led to the conceptualisation of a new multi-faceted capability tool for planning and supporting specialised RDM services in academic libraries.

The weakness, however, is that although this model can be used to understand research outputs in research intensive universities, researchers may have problems in solely depending on this model to understand data curation. This is the case because by trying to strike a balance between factors that enable and impede the creation of research data and its preservation, the model limits the scope of digital preservation. In the words of Ng'eno (2018, p. 28), the CCMF "does not focus in detail on data curation which is a significant constituent in RDM" hence the need to use it in combination with other model(s) as is the case in the current study.

3.6. The Digital Curation Centre Lifecycle Model

The Model was conceptualised "as a generic, curation-specific tool which can be used, in conjunction with relevant standards, to plan curation and preservation activities to different levels of granularity" (Higgins, 2008, p. 134). It was primarily developed to help the UK Digital Curation Centre in achieving three primary roles namely, to train creators of data, data curators, and users; to help individuals and organisations organise their digital resources; and to help organisations plan and implement the preservation of their digital assets. The discrete functions within the DCC Curation Lifecycle Model are conceptually partitioned into three categories: full lifecycle; sequential; and occasional actions. Details of this model are illustrated in Figure 4.

3.6.1. Full-cycle curation activities

Full lifecycle actions encompass the set of actions that need to be performed throughout the lifecycle of digital objects. These are the functions that are performed as long as the digital content is still relevant, even if it exists for ever. There are four full lifecycle actions which will be discussed in this section.

Description and representation of information

This activity involves assigning "administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, to ensure adequate description and control over the long term" (Higgins, 2008, p. 134). To effectively perform this role, Higgins (2008: 134) says curators need to collaborate closely with the data providers to understand the data and assign the best descriptors. In addition, Heidorn (2011) is of the view that curators must

collaborate with other institutions to be able to identify applicable standards and best practices for each data set.

Preservation planning

Since there are various actions and activities within the lifecycle leading to the complete curation of a digital object, Higgins (2008) recommends that there is a need to neatly plan well in advance. On the part of librarians as curators for example, Heidorn (2011) observes that the preservation planning aspect can be embedded in the management and administration of the document objects in the existing digital repository.

Community watch and participation

This activity entails keeping researchers abreast of relevant community activities related to the creation of digital resources. Librarians (as curators) need to keenly follow the new standards, practices and suitable software of primary users of various data objects in particular communities. This is important considering that technologies change at a rapid pace.

Curate and preserve

This activity involves properly managing and taking control of administrative functions to monitor, promote and implement curation and preservation of the digital resources throughout the curation lifecycle.

3.6.2. Sequential actions

Sequential lifecycle actions specify a set of activities that must be undertaken in specific order, to facilitate the curation and preservation process. These actions are eight in total and will be discussion below.

Conceptualise

The activity involves contemplating how the data will be curated including capture and storage facilities. In a university environment, this might start with proposal writing which should highlight well-articulated methods and procedures for the long-term preservation of the research data. This is important because funding agencies are increasingly requiring researchers to include aspects of data management plans (Heidorn, 2011).

Create and receive

Creating data involves producing administrative, descriptive, structural and technical metadata. This also involves creating preservation metadata when data is being created. This is the reason it is recommended that as soon as data collection begins, curators should assist researchers. Receiving or collecting data which is done in line with documented collection policies can be achieved in various ways: collecting data from creators, from archives, from repositories or data centres and, if appropriate, assigning metadata.

Appraise and select

Not all data or digital objects are worth preserving hence the need to appraise and select the ones that qualify for long term preservation. To successfully perform this task, curators need to consult data creators, and make appraisal decisions based on documented institutional policies and legal requirements.

Ingest

At this point, data selected for long term preservation can be transferred to an archive, repository, data centre or other strategic storage facilities by following documented institutional policies and legal requirements.

Preservation action

The curator undertakes actions for long-term preservation while ensuring that its authenticity, reliability, usability and integrity are maintained. During this process, data is cleaned, validated, preservation metadata assigned and acceptable data standards, structure or file formats are assigned.

Store

The data is stored in a secure manner adhering to relevant standards.

Access, use and reuse

This action ensures that users and re-users have access to data and use particular data on a daily basis depending on the restriction access and use conditions imposed by creators.

Transform

The "transform" action involves creating new data from the original by, for example, storing data in different formats, organising data in different sets, or analysing and summarising the data.

3.6.3. Occasional actions

As the term suggests, occasional actions describe those activities that need to be undertaken less frequently as follows: citation needed

Dispose

This involves disposing data that has not been selected for long term preservation as guided by documented policies or legal requirements. Disposition may involve transferring data to a separate archive, repository or data centre or completely destroying it.

Reappraise

Data whose validation has failed may be ear marked for reappraisal and reselection.

Migrate

This involves migrating data to newer file formats that support its continued access and preservation. This is important because hardware or software are not immune to obsolescence.

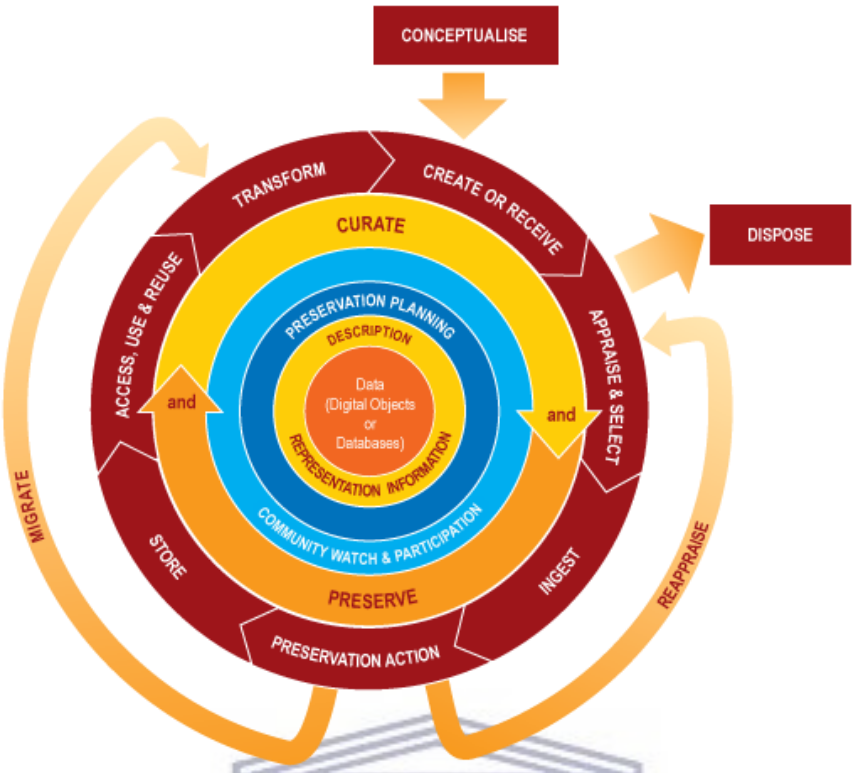


Figure 4. The DCC Curation Lifecycle Model (Higgins, 2008)

3.6.4. Strengths, applications and shortcomings of the theory

One outstanding strength that characterises this model is its customisability to fit into different digital preservation contexts. The model provides a broader picture of data curation (Digital Curation Centre, 2008) and enables researchers, curators and re-users an opportunity to isolate the best practical activities that meet their curation needs. The implication is that curators can add some activities or steps to bridge the missing links or can eliminate some steps or activities in the model that are not appropriate to best realise their curation process. The other notable strength of the model is its ability to provide an implicit hint to the key stakeholders in the curation process whereby, according to Digital Curation Centre (2008), the model demarcates distinct sequential stages that make it easier to identify possible collaborators in the curation process. These collaborators include data creators, data curators and data re-users. The sequential stages also imply that it is easy to document and connect curation policies amongst different stakeholders and that it is easier to identify key tools and services that are required to support curation stakeholders at every level.

In view of the strengths highlighted, a numbers of studies have either used the model individually or jointly with other models to examine and understand RDM. Noting that the

CCMF does not specifically focus on data curation, Ng'eno (2018) used the DCC Curation Lifecycle Model to supplement the CCMF in studying RDM practices in some agricultural research institutes in Kenya. The model proved useful considering that it helped the researcher to understand various activities involved in data curation; these activities include data capture, appraisal, description, preservation, access, reuse and transformation of research data. Brambilla (2015) used the DCC Curation Lifecycle Model to analyse the digital curation practices in academic libraries in Italy by collecting qualitative data from librarians, digital humanists, and data experts. The study revealed a steady headway in the area of data curation; nevertheless, noting some shortfalls in basic IT skills, the study recommended the need for a transformation of LIS curricula in Italy. The DCC Curation Lifecycle Model was used by Heidorn (2011) to determine the skills sets of librarians in the USA in the area of RDM focusing on both full-cycle curation and sequential activities of the Model. Noticing that librarians have greater potential to assimilate knowledge for performing various RDM tasks, Heidorn (2011) challenges librarians to actively in these activities, cautioning that acting to the contrary may force the society to "choose to create a new type of institution to curate digital data" (p. 670). Still in the USA, Shakeri (2013) used the DCC Curation Lifecycle Model to examine data preservation practices among researchers at Kent State University's Liquid Crystal Institute. The study noted that researchers mostly used less dependable free standing devices - computers, hard drives, flash disks and hard discs) to preserve their data.

All these studies (Brambilla, 2015; Constantopoulos et al., 2009; Heidorn, 2011; Ng'eno, 2018; Shakeri, 2013) have verified and validated the appropriateness of this model in implementing data curation activities as well as supporting data curation research studies.

The DCC Curation Lifecycle Model has not been spared from criticism. According to Constantopoulos et al. (2009), the model does not offer guidelines for recording and maintaining various forms of statistics of stored, curated and preserved information as accessed by the users through the queries these users generate. Another notable shortcoming of the model, according to Constantopoulos et al. (2009), is that it does not provide an action of adding new knowledge or combining it to the primary resources or prior knowledge stored in digital repositories. An additional weakness of the model is that it does not provide space for including controlled vocabularies used in different fields of studies which may include geographic names, historical periods, chemical molecules, and biological species (Constantopoulos et al., 2009).The other key limitation of the DCC Curation Lifecycle Model is that is does not incorporate the role of institutional capabilities in data curation which are

paramount in the successful execution of RDM activities. In the words of Ng'eno (2018, p. 34), the model is silent on the aspects of "technical infrastructure, skills and training, collaborative partnerships and legal and policy issues".

3.7.Models adopted for the study

The DCC Curation Lifecycle and CCMF models were adopted for this study. The two models have been chosen because they typically complement and support each other in various aspects and ways.

As already noted earlier, the CCMF was conceptualised to primarily inform and guide various research stakeholders on how to maximise research output in research intensive institutions. This characteristic makes this model an inevitable inclusion in this study because all models reviewed in this chapter place an emphasis on the importance of the creators of digital objects such as research data which are at the centre of data curation activities. This model was also conceptualised to specifically offer guidance on RDM and this characteristic uniquely distinguishes it from other conventional models in the data curation discipline that focusses on general digital objects. Since the focus of the present study is on RDM in universities, this model is therefore, a suitable choice to underpin this study.

As noted earlier in this chapter, the CCMF is characterised by its ability to examine RDM institutional capabilities focusing on attributes of skill and training; technical infrastructure; legal and policy issues; collaborative partnerships; and openness. Since these attributes are investigated in the present study, the CCMF was better placed to understand these perspectives. The study further investigates specific data curation actions which researchers and librarians need to demonstrate when performing RDM processes; the CCMF fails short in demystifying such RDM actions. Observing the gap in the CCMF, the researcher finds it justifiable to adopt a complementary model. The DCC Curation Lifecycle Model is best suited to support and accompany the CCMF in this study because it specifically addresses most key issues necessary for the effective and successful execution of the data curation processes. For example, the DCC Curation Lifecycle Model addresses issues of authenticity and integrity; strategies to provide for adequate knowledge representation and access; support for a predictable preservation lifecycle of assets, as well as attention to the interests of particular communities of practice such as archivists and researchers. Ng'eno (2018) also adopted these models in one study to complement each other in examining and understanding RDM in agricultural research institutes in Kenya.

3.7.1. Mapping the research objectives to the adopted models

Table 5 presents a summary of how the variables gleaned from the two adopted models are tested in this study.

Table 5. Mapping the research objectives to the adopted models

Research objectives	Variables addressed in the study	Sources of variables	Selected prior studies that used these models
To determine research data creation, sharing and re-use practices in public universities in Malawi	Data creation, data formats, data storage facilities, data special features, data re-use, data repurposing, open access, social networks, sources of data, trust of data	CCMF; DCC Lifecycle Model; literature reviewed.	• Shen (2016) used the CCMF to study research data sharing and reuse practices of lecturers at the Virginia Tech in the USA. • Ng'eno (2018) used both CCMF and DCC Curation Lifecycle Model to understand RDM in agricultural research institutes in Kenya • Heidorn (2011) used the DCC Curation Lifecycle Model to analyse the emerging role of libraries in data curation and e-science. • Constantopoulos et al. (2009) reviewed and extended the DCC Curation Lifecycle Model as a key model in guiding curation activities in research and other interdisciplinary domains. • Brambilla (2015) used the DCC Curation Lifecycle Model to review the new roles and professions for digital librarians in Italy.
To investigate research data preservation practices in public universities in Malawi	Metadata, preservation, standards, storage facilities, migration, preservation activities, servers, data formats, data repositories	CCMF; DCC Lifecycle Model; literature reviewed.	
To investigate competencies that librarians and researchers needed to effectively manage research data in public universities in Malawi	Skills, training, competencies.	CCMF; DCC Lifecycle Model; literature reviewed.	
To find out the challenges that affected the management of research data in public universities in Malawi	Challenges, skills, ethical issues, legal issues, funding, rewards and recognition, metadata/data description, data obsolescence, cyberinfrastructure, incentives.	CCMF; DCC Lifecycle Model; literature reviewed.	

3.8. Summary of chapter

Most models reviewed in this chapter are focused on providing a common platform where all data curation activities can be consolidated, coordinated and executed. The models suggest that by systematically merging relevant strands of digital preservation activities into one 'pot', it becomes easier to identify and catalogue problems that constantly stymie digital curation processes. The models were therefore developed to offer guidelines in dealing with problems

that are associated with storage, long-term preservation, maintenance, use and re-use of digital objects such as research data. By thoroughly assessing each of the models, two models namely, CCMF and DCC Curation Lifecycle Model were identified to be most suitable for this study and were thus adopted to support and complement each other.

CHAPTER FOUR

RESEARCH METHODOLOGY

4.1. Introduction

Having presented the research problem and questions in Chapter One which were framed by the literature review and theoretical frameworks discussed in Chapters Two and Three in that order, this chapter focusses on describing and justifying the adopted methodology. Over the decades, various scholars have defined research methodology in a somewhat similar and detailed fashion. However, one of the most cited definitions of research methodology is by Brewerton and Millward (2001) who define a research methodology as a process by which research questions are realised into actions and measured to achieve the overall research aim and objectives. The implication of this definition is that there must be a link between the research questions or objectives, research methodology and data collection methods.

To recap, the aim of the study was to investigate the management of research data in public universities in Malawi with among other aims, to develop best practices for managing research data in these public universities and similar research environments. Four objectives were developed in Chapter One to guide the study as follows:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers needed to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affected the management of research data in public universities in Malawi

This chapter is organised under four interrelated topics namely research paradigms; research design; research methods, reliability and validity, data analysis, ethical consideration and data storage and archiving.

4.2. Research paradigms

A research paradigm refers to key assumptions and beliefs in relation to how researchers view the world; it acts as a framework for shaping the behaviour and perceptions of researchers (Creswell, 2014; Jonker & Pennink, 2010). In most research, philosophical backgrounds are

commonly implicit but they affect the practice and actions of researchers (Wahyuni, 2012). Various authors have independently emphasised the need for researchers to first develop an understanding of research paradigms and choose the one to guide their research because eventually, a paradigm affects how researchers undertake a social study in terms of constructing a social phenomenon (Berry & Otley, 2004; Creswell, 2009; Neuman, 2011; and Saunders, Lewis & Thornhill, 2009). In this spirit, this researcher explores various research paradigms and ultimately, selects one that guides various theoretical assumptions and fundamental beliefs. Four paradigms are identified and widely discussed in the literature. The most common ones include positivism, postpositivism, interpretivism and pragmatism (Wahyuni, 2012, p. 71).

4.2.1. Positivism

Researchers who adopt a positivist paradigm, make standard generalisations which they call nomothetic (Neuman, 2011) by carrying out studies that measure the social phenomenon (Wahyuni, 2012). Their belief is that different researchers investigating the same existing problem will inevitably realise the same and consistent results by performing statistical tests as well as employing similar research methodologies to conduct research in different research settings (Creswell, 2009). Because of the systematic processes involved in investigating a phenomenon, Creswell (2003, p. 12) refers to positivism as the scientific method. One key common aspect amongst positivist researchers is their staunch belief that results realised from one study can be generalised or applied to other populations and across contexts and this philosophical assumption is referred to as naïve realism (Wahyuni, 2012). To them, "positivism is objectivist through and through" (Crotty, 1998, p. 27) and in their research vocabulary, bias or subjectivity does not exist. Creswell (2014) asserts that positivists are behind the conceptualisation of the quantitative research approach and they implement this by conducting experiments or collecting data using questionnaires with closed ended questions which is analysed using statistical software such as SPSS.

4.2.2. Postpositivisim

This group of researchers can best be categorised as 'protestant positivists' who decided to refute some tenets of positivism by declaring that it is not possible to prove the absolute truth of research findings "especially in relation to studying human behaviour in social science" (Wahyuni, 2012, p. 71). Although they buy the idea of research generalisation as promoted by positivist researchers, they admit that knowledge creation is affected by social conditions. They

hold what is commonly called a critical realist stance (Wahyuni, 2012, p. 71) meaning that social reality is primarily structured in a particular context or appropriate law. They believe that social associations fundamentally influence observable phenomena within the social world. For example, they uphold the idea that the values of researchers can affect the outcome of the research findings hence they advocate for the need to validate the findings in order to mitigate possible unforeseen influences (Teddlie & Tashakkori, 2009) or intervening factors. Validating research in this case may entail the choice and use of various methods commonly called triangulation, according to Torrance (2012).

4.2.3. Interpretivism

According to Wahyuni (2012, p. 71), "interpretivists believe that reality is constructed by social actors and people's perceptions of it". They believe that the on-going construction of reality is shaped by the interaction of individuals with diverse backgrounds, assumptions and experiences in the broader social context. They logically assume that the social reality can change owing to the fact that people may have varied perceptions of the social environment surrounding them (Hennink, Hutter & Bailey, 2011). They therefore oppose the concept of objectivity and the idea of absolute truth as championed by positivist researchers. To effectively understand the social world, interpretivist researchers favour direct interaction with individuals they are studying in their social context. Hence they prefer working with qualitative data which they believe provides them rich and detailed information of social reality (Wahyuni, 2012, p. 71).

4.2.4. Pragmatism

This is research paradigm that has shunned the 'paradigm war' being waged between interpretivism and positivist research philosophies (Tashakkori & Teddlie, 1998). Rather than questioning and debating what an ideal research paradigm is, pragmatist supporters direct their energy to identifying the research problem and determining the most suitable research framework. To them, research is considered a continuum, "rather than an option that stands in opposite positions" (Wahyuni, 2012, p. 71) refuting the notion that objectivist and subjectivist philosophies are irreconcilable. It is for this reason that pragmatist researchers believe in a combination of ontology (the perception of researchers towards reality - reality can be either influenced by external social actors or by internal social actors); epistemology (the relationship between the researcher and the subjects – the way the researcher perceives the knowledge generated from the subjects whether acceptable and valid or not); and axiology (concerned with

ethics in dealing with the roles of values in the research process) in investigating a social phenomenon. It is this philosophical view that motivates pragmatist researchers to prefer "working with both quantitative and qualitative data" because in their view, "it enables them to better understand social reality" (Wahyuni, 2012, p. 71).

This study was underpinned by pragmatism. Rather than seeing quantitative and qualitative as competing and contradictory approaches, pragmatism is a basis for mixed methods, seeing qualitative and quantitative research methods as complementary strategies which help researchers to holistically answer the research questions through different means (Creswell, 2014; Silverman, 2011). Unlike interpretivism or positivism which confines the researcher's choice of scientific methods to particular epistemological stances (Denscombe, 2007) , this researcher aligns his philosophical assumptions with those of Saunders et al. (2009) and Ormston, Spencer, Barnard and Snape (2014) who claim that quality in research practice is about having the freedom to select suitable research tools. Creswell (2014, p. 11) declares that adopting a pragmatic paradigm allows the researcher to "use multiple methods, different worldviews, and different assumptions, as different forms of data collection and analysis".

Justification for the choice of the pragmatist philosophy

The researcher draws knowledge and wisdom about the pragmatic paradigm from Ormston et al. (2014) who make it clear that in this research paradigm, the choice of research methods is not confined to any particular known paradigms such as interpretivism or positivism. Pragmatism has been hailed for the suitability and flexibility it brings to the research community. It brings together through collaboration, researchers who are supporters of different paradigms thereby being regarded as a true paradigm for holistically investigating a social reality through its combined use of quantitative and qualitative approaches, and its ability to combine issues at all levels, that is, at macro and micro levels (Onwuegbuzie & Leech, 2005). Creswell (2009) adds that pragmatism allows researchers to consider different world views, assumptions, data collection and analysis techniques. Despite its popularity, there have nevertheless been concerns about pragmatism. Many scholars such as Johnson and Onwuegbuzie (2004) have argued that its proponents have failed to offer a logical and decisive solution to philosophical disputes between interpretivist and positivist paradigms. Its popularity is believed to be a product of two conflicting parties of interpretivists and positivists. For example, Johnson and Onwuegbuzie (2004) observe that pragmatism acts as a bridge between conflicting philosophies and it is therefore enjoying popularity and growth in research

because it benefits from the tenets of both, interpretivist and positivist debates. Nonetheless, these sporadic criticisms were not enough to deter the researcher from using this research paradigm especially considering its well-grounded benefits. The implication of pragmatism is that it affords the researcher more flexibility in selecting appropriate methods that help investigate the problem more holistically. By adopting this paradigm, the researcher was at liberty to employ various methods in terms of choice of research designs, data collection instruments and procedures, and data analysis. In that regard, research designs and tools adopted and described in sections below are selected based on the pragmatic school of thought - allowing the researcher to base his choice of scientific methods on prior scholarly works, adopted models and nature of the research questions.

4.3. Research design: A mixed methods design

A research design is a strategy or a plan for conducting an inquiry (Denzin & Lincoln, 2011; Henn, Weinstein & Foard, 2009). In other words, the design details how data required to answer the research problem will be collected, analysed and validated. Since this research is underpinned by the pragmatic orientation, it was guided by a mixed methods design. Considering that there is sometimes confusion between a research design and a research approach, it has to be first affirmed that mixed methods is a research design, only that commonly, these two terms are used interchangeably. For example, while Creswell (2003), Edmonds and Kennedy (2013), Henning (2004), and Lapan, Quartaroli and Riemer (2012) refer to mixed methods as both a research approach and a research design, Leech and Onwuegbuzie (2009) maintain that mixed methods is a research design in its own right. However, Lapan et al. (2012) admit that a mixed methods research design could be confusing to researchers who are new to mixed methods mainly because there are many types of mixed methods designs from which to choose. Creswell (2014, p. 217) agrees that mixed methods is relatively new in social and human sciences.

By definition, a mixed methods design entails integrating quantitative and qualitative data in one study (Creswell & Creswell, 2018; Heyvaert, Maes & Onghena, 2013). The choice of a mixed methods research design in this study is framed by a pragmatic orientation, which is the basis for all methodological approaches in this research.

4.3.1. Quantitative research

Creswell and Creswell (2018) explain that the key characteristic of quantitative research is that it uses a questionnaire as a data collection instrument which contains closed-ended questions.

Quantitative research is popular in collecting data from large populations through the use of questionnaires which are commended by the research community for being cost-effective and time saving. This research approach is commended for its ability to collect data that researchers can use to predict, explain and validate conceivable relationships among variables (Greene, Kreider & Mayer, 2005; Hair, Bush & Ortinau, 2003; Leedy & Ormrod, 2005). In other words, quantitative research allows researchers to achieve objectivity thereby enabling them to offer reliable explanations about the cause and effect between or among variables being tested. Over the years, researchers have come to agree that results realised through quantitative research are mainly free from human subjectivity because, according to Patton (2002) and Durrheim and Painter (2006), it gathers data which is systematic and standardised hence its findings can be generalised to a population of a similar nature. Apart from its ability to achieve objectivity, quantitative research is hailed for its simplicity in analysing the data it collects. However, this research approach is not short of limitations which have exposed it to criticism from its opponents. Quantitative research is described as weak in directly soliciting the actual voices of the participants being studied thereby failing to provide an adequate understanding of the "context or setting in which people talk" (Creswell & Plano Clark, 2011, p. 12).

4.3.2. Qualitative research

Qualitative research is characterised by the use of open-ended questions that guide the researcher in conducting interviews with respondents (Creswell & Creswell, 2018). Open-ended questions are also common in gathering qualitative data from questionnaires. In addition to interviews, qualitative data can be gathered by studying artefacts such as dairies, journals, and reports, to mention a few. Qualitative research is hailed for its ability to collect data about behaviours, experiences, attitudes, beliefs and thoughts of people (Patton, 1990, p. 22). Qualitative research has also received criticism from its challengers based on its limitations. Results from qualitative research can be influenced by a researcher's personal interpretations leading to bias (Creswell & Plano Clark, 2011, p. 12). In addition, the fact that qualitative inquiry focusses on a particular context or setting implies that its resultant findings cannot be generalised "to a large group because of the limited number of participants studied" (Creswell & Plano Clark, 2011, p. 12).

4.3.3. Implementation of mixed methods design in the study

World leading scholars in mixed methods research (Creswell & Plano Clark, 2007; Leech & Onwuegbuzie, 2009; Morgan, 2007; Teddlie & Tashakkori, 2009) have independently

sanctioned pragmatism as a base for mixed methods which this study embraces. Creswell and Creswell (2018) report that qualitative and quantitative designs have each inherent weaknesses and biases. The idea of mixed methods was therefore conceptualised to deal with the flaws associated with each research design. Mixed methods design draws its features from both quantitative and qualitative research designs. Creswell and Creswell (2018) determines three key characteristics of mixed methods research, namely it uses predetermined and emerging methods; it integrates open and closed ended questions, collects multiple forms of data using all conceivable possibilities; and it bases data analysis on statistics and texts.

Creswell and Plano Clark (2011, p. 12) highlight some notable advantage of a mixed methods research design. They report that historically and practically, one of the most recognised advantages is its ability to offset the limitations of both qualitative and quantitative research. They also point out that by not tying researchers to particular data collection methods, mixed methods design grants researchers more evidence to answer a research problem which cannot be answered by either a quantitative or qualitative approach alone. Mixed methods research offers researchers flexibility in using "multiple worldviews or paradigms" (Creswell & Plano Clark, 2011, p. 13) – typically in reference to the pragmatic paradigm.

The implication of mixed methods in the current study is that it granted this researcher an opportunity to adopt and combine multiple methods in one study. Specifically, the study collected data from three categories of participants (researchers, librarians and directors of research); used questionnaires and interviews as data collection instruments; and used descriptive statistics and themes to analyse and report the findings. While researchers are at liberty to integrate mixed methods at different levels - design, methods and interpretation and reporting (Fetters, Curry, & Creswell, 2013, p. 2134), combination in this study was implemented at methods and interpretation and reporting levels.

Creswell and Plano Clark (2011, p. 70) and Creswell (2014, p. 15) identify six mixed methods research designs from which researchers can choose. They include convergent parallel mixed methods, explanatory sequential mixed methods, exploratory sequential mixed methods; embedded design, transformative design and multiphase mixed methods designs. This study was guided by a convergent mixed methods approach (also referred to as concurrent design).

Convergent parallel mixed method design

Convergent mixed methods is commonly referred to as the convergent design (Creswell, 2014, p. 15; Creswell & Plano Clark, 2011, p. 70). It is characterised as a research approach where

the researcher merges the study phases with an intent to compare quantitative and qualitative data which are both collected and analysed within a similar time frame (Creswell, 2014, p. 15; Creswell & Plano Clark, 2011, p. 70; Fetters et al., 2013, p. 2134). Creswell and Plano Clark (2011, p. 70) explain that at data collection and analysis stages, the researcher independently but equally dedicates priorities to both approaches (qualitative and quantitative) "and then mixes the results during the overall interpretation". Its main purpose, according to Creswell and Plano Clark (2011, p. 70), is to obtain different but complementary data on the same topic so as to understand the research problem holistically.

In this study, distribution of questionnaires to researchers and librarians and conducting of interviews with directors of research was carried out concurrently. Like other mixed methods designs, convergent mixed methods is the basis for complex mixed methods strategies. By collecting both qualitative and quantitative data, this researcher was able to holistically investigate research data management practices in public universities in Malawi. To achieve this, the two forms of data (qualitative and quantitative) were collected almost simultaneously, analysed separately but a combination of these sets of data were implemented at the analysis and reporting stage (Creswell & Creswell, 2018) "to see if the findings confirm or disconfirm each other" (Creswell, 2014, p. 19).

The convergent design has three key advantages according to Creswell and Plano Clark (2011, p. 70). It is an appealing design to researchers new to mixed methods research because it "makes intuitive sense" (Creswell & Plano Clark, 2011, p. 70) implying that naturally, it is theoretically and practically logical to understand and to implement. It is an efficient design because it affords the researcher an opportunity to collect both qualitative and quantitative data at one phase and at almost same time. The fact that both forms of data are analysed separately using their natural techniques, implies that it enriches research collaboration in the sense that the research team can "include individuals with both quantitative and qualitative expertise" (Creswell & Plano Clark, 2011, p. 78). This researcher benefited from the first two advantages.

A decision to adopt a mixed methods design in this study was arrived at in consideration of the nature of data required to comprehensively answer the research question. The researcher hoped that quantitative and qualitative data was both important in exploring the RDM practices in public universities in Malawi from the perspectives of librarians, researchers and directors of research. The literature shows that debates about quantitative versus qualitative research as discrete research approaches have blurred. Instead, many scholars increasingly agree that

studies can be more quantitative or more qualitative in nature hence the emergence of mixed methods research (Creswell, 2014) whose backbone is a pragmatic orientation. To investigate RDM practices at the two universities, the study adopts a predominantly quantitative approach supplemented by a qualitative approach.

In order to collect quantitative data from a larger population (researchers and librarians), a questionnaire was distributed to the respondents. The two research sites are geographically separated by a distance of about 800 kilometres hence the choice of a questionnaire which is hailed to collect data from populations which belong to different geographical regions (Schutt, 2006). In addition, by adopting quantitative research, the researcher was able to easily statistically analyse data (Creswell, 2009) which was collected from 187 researchers and 36 academic librarians.

Since qualitative research is applauded for its capability in collecting a corpus of rich and detailed data (Patton, 1990, p. 22) through conducting interviews with directors of research, this researcher was able to collect detailed information about these participants' experiences, attitudes, beliefs and thoughts about research data management practices in Malawian public universities.

In implementing this mixed methods design, data collected from librarians and researchers (predominantly quantitative) and from directors of research (qualitative) was analysed separately and merged at the reporting phase (Luck, Jackson & Usher, 2006). Qualitative and quantitative data was used to validate each other and formed the basis for discussions and conclusions regarding the RDM practices in public universities in Malawi. This process is called triangulation (McMillan, 2004; Patton, 2002) and it has been defined by the scholarly community as the use of multiple data collection methods (Wahyuni, 2012, p. 71) in order to gather credible and dependable information (Decrop, 1999) which guards against limitations inherent to any single research method (Creswell & Creswell, 2018; Williamson, 2005).

This means shortcomings of quantitative data collected from researchers and librarians were offset by qualitative data collected from directors of research and vice versa, hence a justification for the adoption of a mixed methods design in this study. For example, quantitatively studying researchers and librarians meant it was possible to generalise the results university wide but it was going to be difficult to adequately understand an individual researcher or librarian (Creswell & Plano Clark, 2011) in terms of RDM practices. Likewise, in line with Creswell and Plano Clark's (2011) observation, qualitatively studying directors of

research in this study meant that despite the study achieving an individual understanding of a particular participant, it was going to be impossible to generalise the results university wide. In this spirit, adopting a mixed methods design was inevitable. In summary, triangulation helped the researcher to achieve consistency by cross-checking data from the two sources "in order to enhance the robustness of findings" (Wahyuni, 2012, p. 71) and "to build a coherent justification for the themes" (Creswell, 2014, p. 201).

4.4. Research methods

Since the study is inspired by the pragmatic school of thought, both, quantitative and qualitative research approaches are adopted. Qualitative research provides flexible ways in which to collect, analyse and interpret data; it provides a holistic view of the phenomenon under investigation in addition to allowing the researcher to interact with the research participants (Creswell, 2009; Matveev, 2002). On the other hand, Creswell (2009) notes that data collected through quantitative methods provide information which can easily be analysed to statistically generalise respondents' explicit and implicit claims. It is the complementary strengths of these stances that support the present study's philosophical stance of pragmatism which is the basis for mixed methods research methodology. As already mentioned, it is clear that this study conforms to the tenets of a mixed methods design by collecting data using two data collection instruments and procedures and from multiple sources. The researcher agrees with Creswell and Plano Clark (2011, p. 12) and McMillan (2004) who note that through triangulation, qualitative and quantitative data is collected almost simultaneously to take advantage of the strengths of either method and at the same time to offset the weaknesses of the other. By triangulating data collected using questionnaires and interviews, the researcher is convinced that the study yielded credible and most convincing findings and conclusions.

4.4.1. Study population

According to Sekaran and Bougie (2010), a study population refers to the entire group of people, events or things of interest that the researcher wants to investigate. The population of this study involves lecturers (researchers), library staff and directors of research in public universities in Malawi. This study was conducted at two universities in Malawi which for ethical reasons have been given the pseudonyms of University 1 (UNI1) and University 2 (UNI2). Details about universities in Malawi are provided in Chapter One (see section 7.1).

As can be seen in Table 6, UNI1 has six faculties with a total number of 23 departments. On the other hand, UNI2 has three faculties with 15 departments. UNI1 can be described as an

interdisciplinary university considering that it focuses on diverse areas of specialisations that are offered under the faculties of Humanities and Social Sciences; Health Sciences; Environmental Sciences; Science, Technology and Innovation; Tourism & Hospitality Management; and Education. On the other hand, UNI2 offers pure health related disciplines which are offered through the faculties of Biomedical Science and Health Professions; Medicine; and Public Health and Family Medicine.

Table 6. Distribution of universities under study by faculties and departments

Universities			
UNI1		UNI2	
Faculty	**Department**	**Faculty**	**Department**
Humanities & Social Sciences	Information Sciences	Biomedical Science and Health Professions	Pathology
	Governance Peace & Security Studies		Pharmacy
	History & Heritage Studies		Physiotherapy
	Theology & Religious Studies		Medical Laboratory Sciences
	Communication Studies		Biomedical Sciences
	Languages and Creative Arts	Medicine	Surgery
Health Sciences	Nursing & Midwifery		Medicine
	Optometry		Anesthesia
	Biomedical Sciences		Pediatrics,
Environmental Sciences	Forestry & Environmental Management		Gynecology
	Fisheries & Aquatic Sciences		Ophthalmology
	Built Environment	Public Health and Family Medicine	Public Health
	Water Resources Management & Development		Family Medicine
	Agrisciences		Health Systems and Policy
	Geosciences		Community Health.
Science, Technology and Innovation	Biological Sciences		
	Chemistry		
	Energy Systems		
	ICT		
	Mathematics & Statistics		
	Physics & Electronics		
Education	Education Foundations & Teaching		
	Learning and Curriculum Studies		
Tourism & Hospitality Management	Hospitality Management		
	Tourism		

Source: UNI1 (2017) and UNI2 (2018)

Both universities have libraries which serve the academic community with various information services. Current and relevant information sources are particularly vital if lecturers, researchers and students are to keep abreast with emerging issues in their areas of practice. UNI1 has a library that supports the teaching, learning and research activities of students, lecturers, and non-academic members of staff, and members of the surrounding community. Over the years, the library at UNI1 has invested in print materials and electronic resources and information and communication technologies. It has a collection of 53,000 books, 68 desktops computers, 403

reading chairs, 62 reading tables, 111 shelves, three heavy duty photocopiers, eight printers (Chawinga & Majawa, 2018). Likewise, UNI2 supports the academic programmes of the college (UNI2, 2018). In order to achieve this goal, the library has more than 30,000 volumes of books and subscribes to a number of electronic databases which offer over 20,000 full text electronic journal articles.

4.4.2. Sampling

Onwuegbuzie and Collins (2007, p. 281) acknowledge that making sampling decisions in mixed methods research like the current one is a complicated process because the researcher must design sampling schemes for both, qualitative and quantitative research components. Many researchers have indicated that traditionally, random sampling is associated with quantitative research whereas non-random sampling is naturally linked to qualitative research (Creswell & Plano Clark, 2011; Onwuegbuzie & Leech, 2005; Teddlie & Yu, 2007). In that regard, in this study, the researcher applies these sampling techniques in line with their traditional research approaches. Specifically, purposive sampling is used for qualitative research and a census (a non-sampling technique) is used for quantitative research.

4.4.2.1. Purposive sampling: Selection of research sites and directors of research

The study used purposive sampling for two scenarios. First, to select the research sites and secondly, to select directors of research. This sampling technique has been defined by Teddlie and Yu (2007, p. 77) as "selecting units (e.g., individuals, groups of individuals, institutions) based on specific purposes associated with answering a research study's questions". In other words, it is employed where the researcher's intention (due to the nature of the research problem) is to obtain information from a particular setting or individuals that cannot be obtained from other choices. It focusses on gathering rich data from individuals who have experienced the phenomenon. Because purposive sampling is inherently associated with qualitative research, a range of its sampling techniques "have also been referred to as nonprobability sampling or purposeful sampling or qualitative sampling" (Teddlie & Yu, 2007, p. 80). According to Babbie (2004, p. 94), purposive sampling involves the selection of units to be observed on the basis of the researcher's own judgment about which ones will be the most useful. In this study, purposive sampling was exercised because it was necessary to ensure that information is obtained from the sample that comprises information rich (Onwuegbuzie & Leech, 2007; Patton, 1990) participants who are directly affected by RDM issues (Moustakas, 1994).

Selection of research sites

As already indicated elsewhere in this document, purposive sampling was used to select the two research sites. To begin with, the two universities were selected purposively by considering two factors. First, the researcher wanted public universities which commanded a good research record. These two universities have existed for over two decades and have been involved in various research activities. The other two public universities which were only established after 2012, did not match this first selection criterion, hence they were not selected. Second, the researcher wanted to make the study more interdisciplinary by drawing data from researchers with diverse educational backgrounds and expertise, particularly in health sciences and diverse academic disciplines. Based on this premise, UNI2 was selected because it focusses on health related research while UNI1 was selected because it offers diverse disciplines.

Selection of directors of research

These are the heads of the research units and are part of the senior university management team. They were included in the study because they were in a position to answer questions on the universities' involvement in scaling up research output, RDM and challenges. This is acceptable because Creswell and Plano Clark (2011, p. 173) say "in qualitative research the inquirer purposively selects individuals and sites that can provide the necessary information".

4.4.2.2. Census: Identification of lecturers and librarians

This study made use of a census in identifying lecturers and librarians meaning that no sampling method was used as all lecturers and librarians were asked to participate in the study. According to Israel (2013), a census is a technique that involves the study of all elements that form the whole study population. This technique is generally feasible for small populations (Israel, 2013) although Cantwel (2008, p. 90) says it is also popular for large populations such as a national population. Cantwel (2008, p. 90) likens a census to a survey in the sense that both use questionnaires, similar data collection procedures and similar data processing and analysis. The only difference between a survey and a census is that "[U]nlike a sample survey, in which only a subset of the elements is selected for inclusion and enumeration, a census generally does not suffer from sampling error" (Cantwel, 2008, p. 90). It generally collects data on all eligible elements in a defined population (Cantwel, 2008, p. 90). In this study, a justification for using a census in based on a recommendation by Israel (2013) who says that, where the population is less than 200 in a particular study site, the whole population should be studied.

Lecturers

In Malawian public universities, a minimum qualification for a lecturer position is a master's degree. Lecturers understand both the principles of research and are able to conduct research. This premise is based on the understanding that most lecturers obtain a master's degrees after doing some research as part of a master's course; rarely do universities offer masters programmes that exclude the component of research that culminates in a thesis or dissertation. The data provided by UNI1 shows that it has a population of 147 lecturers whereas UNI2 has a population of 130 lecturers. Building on the work of Israel (2013), since the population for lecturers at each university was less than 200, the researcher decided to include all lecturers. See Table 7 for details about these statistics.

Library staff

In Malawian university libraries, library staff involved in helping researchers as users of the library are those qualified with a certificate in Library and Information Science or higher. Since information gathered from UNI1 and UNI2 showed that each had a population of less than 200 library staff, the researcher included all of them in the study (see Israel, 2013). See Table 7.

Thus, based on a purposive sampling technique and census method as discussed in the preceding section, in all, 277 lecturers, 38 library staff and two directors of research were included in the study giving a total sample size of 317. See Table 7.

Table 7. Sample size (*n*= 317)

University	**Category of population**		
	Lecturers	Library staff	Directors of research
UNI2	130	18	1
UNI1	147	20	1
Sub-totals	**277**	**38**	**2**
Total sample			**317**

Source: UNI2 (2017) and UNI1 (2017)

4.4.3. Data collection instruments

4.4.3.1. Questionnaires

In order to gather quantitative data, the researcher used a questionnaire which is defined by Babbie (2004) and Connaway and Powell (2004, p. 146) as a document containing a set of

questions expected to be completed personally by the respondents. Questionnaires are said to be the most popular form of data collection instrument "in any research involving human subjects" (Pickard, 2007, p. 183). One strength of questionnaires is that they are capable of collecting numerical data provided by respondents without the presence or influence of researchers (Blaxter, Hughes & Tight, 2006).

Primarily, questionnaires can be categorised into two broad forms in terms of how the actual data collection is executed: self-completed and interviewer completed (Saunders, Lewis & Thornhill, 2012). Self-completed questionnaires are completed by the respondents and depending on the distribution strategy, they can be further categorised as web-based or Internet mediated questionnaires, intranet mediated questionnaires, postal or mail questionnaires and delivery and collection questionnaires (Saunders et al., 2012). On the other hand, interviewer-completed questionnaires are completed by the researcher by reading the questions to the respondents and the interviewer records the answers chosen by the respondent; depending on the communication mode between the interviewer and the respondent, these questionnaires can be categorised as telephonic questionnaires or structured interviews (Saunders et al., 2012). In this study, the researcher used the delivery and collection method whereby the researcher hand-delivered printed questionnaires to librarians and lecturers and later collected them after they were completed by the respondents. This approach was favoured because the researcher wanted to allow respondents to fill in the questionnaires at a convenient and appropriate time to them without the influence of the researcher (Blaxter et al., 2006). The researcher used the hand delivered questionnaire other than web-based questionnaires because the researcher wanted to maximise the response rate. Researchers are busy people and the researcher anticipated that some researchers could find it tedious and inconveniencing filling the questionnaire online, which could as well require them having access to the Internet which can sometimes be unreliable in Malawi.

The questionnaire items (see Appendix A for library staff and Appendix B for researchers) were largely informed by the two models (CCMF and DCC Curation Lifecycle Model), the research questions and the literature. The researcher employed six research assistants to help with administering the questionnaire and its collection.

Researchers need to consider two types of questionnaire questions, according to Denscombe (2007). They include open ended questions which afford the respondents to decide on the wording and length of the answer allowing them to express their complex views, and closed

ended questions whose answers are predetermined in advance by the researcher and respondents are only required to select answers from the options provided. This study was predominantly quantitative in nature. The study required respondents to provide specific information as it was informed by the adopted models and related literature. In that regard, this researcher specified the answers from which respondents selected the best options that applied to them implying that the questionnaires contained closed ended questions. Two forms of closed ended questions were asked in the questionnaire including yes or no questions while other questions were in form of a Likert scale or some form of quantitative measure. The use of standardised questions and answers in this study allowed this researcher to interpret the data the same way by all respondents (Saunders et al., 2012). In instances where respondents were unable to select answers from the provided options, they were asked to mention their own answers through the provision of the option of "other". Being quantitative in nature, it was also easier to analyse the data using SPSS and afforded a quick and easier way to interpret the findings. The questionnaire has been used by similar prior studies as can be seen in Table 8.

The key challenge associated with a questionnaire is that it may lead to low response rates in situations where respondents are not willing to answer the questionnaire. To maximise the response rate, gentle weekly reminders were made to the participants until a reasonable number of questionnaires was achieved. Another drawback of a questionnaire is that where respondents are not competent to answer the questions due poor understanding of the concepts being investigated, it may result into biased, inaccurate responses, and incomplete responses (Babbie, 2004). In this study however, the study collected data from researchers and librarians who had some basic understanding of RDM; this observation is based on the premise that researchers conduct research as part of their core duties while librarians serve researchers with information research services. In addition, when distributing the questionnaire, the researcher explained and clarified to respondents the concept of RDM. To eliminate elements of possible biasness resulting from the natural weakness of a questionnaire, the researcher triangulated the data collected using a questionnaire with qualitative data collected through interviews.

4.4.3.2. In-depth interviews

Many researchers have defined research interviews in a very similar fashion by using different terminologies or words but maintaining the same meaning. One of the most cited definitions of the research interview is provided by Saunders et al. (2012, p. 372) who refer to it "as a purposeful conversation between two or more people requiring the interviewer to establish

rapport, to ask concise and unambiguous questions to which the interviewee is willing to respond and to listen attentively". Interviews are extensively used by researchers in the social sciences discipline as a tool for collecting detailed information concerning a topic or subject (Wilkinson & Birmingham, 2003). Interviews are mostly feasible when the researcher wants to collect qualitative data characterised by Pickard (2007) as mainly descriptive, in-depth data whose nature is too complicated to collect using other forms of data collection instruments such as questionnaires.

Three distinct categories of interviews are identified by Saunders et al. (2012). They include structured; semi-structured and unstructured or in-depth interviews. In structured interviews, an interviewer asks respondents a series of predetermined questions that contain a limited set of responses (Pickard, 2007). Semi-structured interviews are where the researcher prepares a list of themes broken into some key questions which guide the researcher in asking the respondent (Saunders et al., 2012). More precisely, Saunders et al. (2009) describe these types of interviews as a hybrid in nature considering that they draw their characteristics from both structured and in-depth interviews. They use predetermined themes and questions that guide the researcher when conducting an interview, "while keeping enough flexibility to enable the interviewee to talk freely about any topic raised during the interview" (Wahyuni, 2012, p. 74). Finally, unstructured or in-depth interviews entail the use of open ended questions that afford the interviewees the opportunity to express their opinions, feelings and thoughts about the research problem (Saunders et al., 2012) allowing interviewees to narrate their story in their own words and in their own setting. According to Wahyuni (2012, p. 74), "the use of an in-depth qualitative interview is considered as the appropriate format for case study research because in-depth questions cannot be answered briefly". One major advantage of interviews is that they enable the researcher to interact directly with participants thereby providing new insights about the issue being researched (Russ-Eft & Preskill, 2001) and unexpected variables may emerge (Muijs, 2010, p. 8). The researcher is justified in using interviews because Creswell (2014) and McNamara (1999) argue that the most important sources of qualitative research are interviews.

This study employed in-depth interviews to collect data from directors of research of the two universities who provided rich data about specific issues in relation to the RDM in their respective universities. During the interviews, the researcher was able to ask for more explanations and examples on the responses provided by participants to gain a deeper understanding of the RDM issues in the two public universities. An interview guide was

developed and used to guide the researcher during the interviews (see Appendix C). By drawing lessons from Rubin and Rubin (2005), the interview guide was structured in such a way that it contained "open-ended main questions, follow-up questions and probes" (Wahyuni, 2012, p. 74). Follow-up questions were used to "explore the particular themes, concepts, ideas and unexpected thoughts" (Wahyuni, 2012, p. 71) about RDM practices. Probes which were prepared in advance, were used to keep the discussion flowing in addition "to clarify some discussion points by asking for more details or examples of what had been said" (Wahyuni, 2012, p. 71) by the directors of research.

In line with Kvale and Brinkmann's (2009) recommendation, the interview guide covered the purpose of the interview, key questions and a conclusion. The interview guide played a role in making sure the researcher was consistent between the interviewees (Boyce & Neale, 2006). The researcher decided to conduct interviews with only directors of research because of one key factor, that is, interviews are time consuming to conduct and to analyse (Barker, Pistrang & Elliott, 2005). By using in-depth interviews, the researcher was able to holistically gain an understanding of the feelings, thoughts and experience (Pickard, 2007) of each director of research about the RDM practices in the respective universities. In that regard, the researcher was able to collect data that represented true opinions, feelings, emotions and experiences of the respondents; such an action resonates with Denscombe (2007) who guarantees that the nature of in-depth interviews allows researchers to explore the research problem in depth and in detail, rather than just reporting short answers in a word or two. More importantly, pragmatic or mixed methods research, where interviews are used to collect data, helps researchers to better understand lived experiences or reality in social context (Sandberg, 2005) and "what the world means to the person or group being studied … in the social sciences" (Willis, 2007, p. 6).

Conducting interviews enabled the researcher to interact with key players (directors of research) in the research process in public universities. These participants provided important data which was relevant and inevitable towards the completion of this research (Denscombe, 2007). Although interviews are hailed for their ability to collect rich qualitative data in social sciences (Wilkinson & Birmingham, 2003), they are not short of shortcomings. According to Wilkinson and Birmingham (2003), interviews are not feasible in collecting data from a large number of respondents. For this study, interviews were conducted with two directors of research from each university. Another challenge posed by the use of questionnaire is the time they take to complete a single interview session (Wilkinson & Birmingham, 2003). In the

present study, while ensuring that the questions asked gathered rich, relevant, and detailed data, a reasonable number of questions were asked to ensure that one interview session should not exceed 60 minutes. The literature shows that other related studies used interviews in investigating RDM as can be seen in Table 8.

4.4.4. Data collection procedures

Two forms of data collection tools were designed: questionnaires which were used to collect data from librarians and researchers; and an interview schedule which was used to collect data from directors of research. Furthermore, before the data collection exercise commenced, the researcher had to address ethical clearance issues which involved seeking and being granted permission by university authorities to conduct the study in the two purposively selected universities.

The researcher administered the questionnaire with the help of six research assistants. Respondents were given a period of one week to complete the questionnaires which were later collected by the researcher and the research assistants. Before distribution of the questionnaires commenced, the researcher visited the registrar's offices of both universities which issued circulars to all lecturers via the universities' mailing list advising them that the researcher will be collecting data from them. Further, an introductory letter was issued to the researcher and research assistants by the two offices which were presented to respondents. Where the respondents were not available in offices, questionnaires were left with the secretaries who delivered them to the lecturers and completed questionnaires were collected via the same secretaries. The researcher trained the research assistants in data collection especially in ensuring that they observed and adhered to ethical issues. The research assistants were trained how to ensure that throughout the study, they strictly adhered to the University of the Western Cape (UWC)'s Code of Conduct for Research.

Before resorting to conducting the interviews, the researcher made appointments with the two directors of research. Both directors of research confirmed the day and time they were available for the interviews. After seeking permission to have the discussion recorded, the researcher used a Samsung Smart Phone inbuilt recorder. During the process of the interviews, the researcher also took some notes as backups to the audio recording. Taking notes during interviews has been reiterated as important because Wahyuni (2012, p. 74) observes that "besides recording the interview, the researcher should also take notes during and soon after each interview to record additional information in the form of research memos". Audio

recording interviews has become easier with the advancement and proliferation of ICTs such as mobile technology applications. The advantage of audio recording the interviews, according to Denscombe (2007), is that, aside from ensuring permanency of the recorded conversation, it lends itself to being checked and verified by other researchers thereby minimising scepticism about the collected data. As recommended by Kvale and Brinkmann (2009), the researcher conducted a briefing before the interview commenced explaining the aim of the interview and emphasising the confidentiality, anonymity and the voluntary nature of the interview and a debriefing afterwards which focused on asking the researchers their general comments on the interview and thanking them for taking part in the study. During debriefing, the researcher encouraged the participants to "ask questions, make comments or add any information that was not discussed during the interview" (Wahyuni, 2012, p. 74).

4.5. Reliability and validity

Positivist researchers propagate that "research should rely heavily on reliability and validity to ensure its replicability and generalisability" (Wahyuni, 2012, p. 74). Reliability is concerned with consistency, predictability and stability of instruments used in research (Kumar, 2010, p. 181). It questions whether the same results could be recorded if another researcher were to conduct similar research using the same instruments. On the other hand, validity establishes appropriateness and accuracy of the research procedures used to find answers (Kumar, 2010, p. 177). Validity questions if the research is really measuring what it is supposed to measure, thus getting answers to questions it was intended to answer.

In this study, to ensure validity and reliability of the quantitative data, the questionnaire content was rigorously perfected. The researcher agrees with Creswell (2014) that an instrument should be subjected to criticism by a diverse groups of people such as peers, academics, researchers and feedback from seminars and conferences. Thus, the researcher submitted the questionnaire to experts in the fields of RDM, data curation and library and information science to determine its appropriateness, and whether the research questions could help answer the research problem.

Golafshani (2003, p.598) identifies three types of reliability referred to in quantitative research: first, the "degree to which a measurement, given repeatedly, remains the same", second, the "stability of a measurement over time"; and third, the "similarity of measurements within a given time period". To achieve reliability in this study, the Cronbach's alpha (α) was applied to each of the questionnaire items to determine the degree of consistency. According to Tavakol

and Dennick (2011, p.53), Cronbach's alpha provides a "measure of the internal consistency of a test or scale; it is expressed as a number between 0 [zero] and 1 [one]". Test scores that range between 0.7 and 1.0 are acceptable because they indicate that the data is reliable. Generally, a test score closer to 1(one) is considered to be of high reliability and acceptable while a test score closer to 0 (zero) represents poor reliability and it is less acceptable (Tavakol & Dennick, 2011). The researcher worked out Cronbach's alpha of each questionnaire question and all questions that did not surpass the minimum threshold of 0.7 were perfected and tested and piloted until the minimum 0.7 was achieved.

For qualitative research, generalisability "has been the major point of criticism of qualitative research" (Wahyuni, 2012, p. 77). The application of reliability and validity concepts in qualitative research has been problematic because qualitative research does not subscribe to the tenets (missions and agendas) of positivist research for which the concepts of reliability and validity were specifically developed (Parker, 2003). Because its mission and agenda is to generate context based knowledge which is hugely characterised by its uniqueness other than generalisability, Kalof, Dan and Dietz (2008) and Bryman (2012) advise that reliability and validity are not suited to reinforce the reliability of results in qualitative research. Consequently, issues of reliability and validity in qualitative data are substituted by quality, rigour and trustworthiness (Davies & Dodd, 2002; Golafshani, 2003; Wahyuni, 2012, p. 74; Stenbacka, 2001). Creswell and Miller (2000), Golafshani (2003), Healy and Perry (2000), and Patton (2002) endorse triangulation as typically a strategy for improving quality and trustworthiness in qualitative research. Patton (2002) argues that triangulation which involves combining quantitative and qualitative approaches is dealt with in two stages of the study - at the study design and analysis of results stages. At the research design stage, this study adopts both qualitative and quantities designs and at the analysis of results stage, trustworthiness, rigor and quality, conclusions were based on findings from the two adopted research approaches i.e. data from the questionnaires and interviews formed the basis for discussing the findings.

4.6. Data analysis

Data analysis is the stage at which the gathered data is transformed into information (Mouton, 2001, p. 108). Primarily, it is the process of drawing inferences from raw data. Consistent with the pragmatic paradigm which allowed the study to be a multi-methods (Wahyuni, 2012, p. 74) or methodological triangulation (Patton, 2002), the study involved analysing both, data

collected using a questionnaire (quantitative) and data collected using an interview guide (qualitative).

Before the data analysis commenced, the researcher performed three important activities: "data storage, transcribing audio sources [for qualitative data], and cleaning the data" (Wahyuni, 2012, p. 74). This is important because Boeije (2010) recommends that raw data needs to be managed or prepared before it can be analysed. In terms of data storage, the researcher had copies of collected data (questionnaires and field notes) stored in a locked filing cabinet and SPSS data sets and audio recordings were stored on the researcher's password-protected computer. Transcribing recorded interviews involves transforming the recorded interview into texts. Although Wahyuni (2012, p. 74) says this activity can be outsourced, the researcher performed this activity himself owing to his experience in conducting qualitative research. Transcribing focussed on extracting content leaving out actual expressions, a practice which Oliver, Serovich & Mason (2005) refer to as denaturalised transcription. Cleaning data helped this researcher in dealing with concerns of anonymity and confidentiality by removing all footprints that participants had left behind (intentionally or unintentionally) in the process of completing the questionnaires and answering the research interview questions. For example, some respondents mentioned the universities from where they came yet universities were anonymised; such information was removed as part of the cleaning process.

4.6.1. Quantitative data

For quantitative data, the SPSS version 20 was used to capture raw data and to execute the descriptive statistics in form of percentages and frequencies. The researcher used the independent samples t-test and one-way analysis of variance (ANOVA) in the SPSS to analyse the data. Specifically, the researcher used percentages, mean, independent samples t-test and one-way analysis of variance (ANOVA) to analyse the data.

4.6.2. Qualitative data

The researcher analysed qualitative data using a thematic analysis approach which is a method for identifying, analysing, and reporting patterns (themes) within data (Braun & Clarke, 2006; Given, 2008). A thematic analysis approach is also commonly called qualitative content analysis (Wahyuni, 2012). The researcher transcribed the interview data (see, Braun & Clarke, 2006) from which analysis was carried out by combining and cataloguing related patterns into themes (Aronson, 1995, p. 3; Boeije, 2010). To achieve this, the researcher used the research

aims and questions as guides in "the process of cutting the collected texts into pieces and logically recombining them" (Wahyuni, 2012, p. 74).

In reporting the findings, the researcher used both direct quotes and paraphrased common ideas emerging from participants. This means that whereas in qualitative research the researcher concentrated on noting and reporting the social reality resulting from the field data (Silverman 2011), in quantitative data, the researcher concentrated on testing the variables and establishing their relationships (Creswell, 2014). Triangulation in this context provided 'checks and balances' to each form of data especially considering that opponents of qualitative research (positivism) criticise it as lacking generalisability while counterparts of quantitative research (interpretivism) criticise it as lacking the ability to collect rich and detailed data. Analysis of the data collected in this study is based on the convergent mixed methods design (Creswell, 2014, p. 220; Creswell & Plano Clark, 2011, p. 69). A summary of how this exercise was implemented in this study is presented in Figure 5.

While acknowledging the existence of challenges in merging qualitative and quantitative in convergent mixed methods designs, Creswell (2014, p. 222) mentions various ways the sets of data can be merged in this design. They include a side-by-side approach, data transformation and joint display of data. The last two are beyond the scope of this study and are not discussed. However, of interest in this study is the side-by-side analysis approach which this study embraces. In the side-by-side approach, comparisons of data are embedded in the data analysis and presentation section meaning the researcher can choose to present quantitative first followed by qualitative themes that support or disconfirm quantitative data. Alternatively, the researcher may present qualitative findings first followed by quantitative statistical results, says Creswell (2014). In this study the researcher presented the quantitative results first followed by qualitative data which was used to compare and validate the quantitative results. The researcher presented quantitative data first because the study was predominantly quantitative (Creswell, 2014) complimented by the qualitative approach.

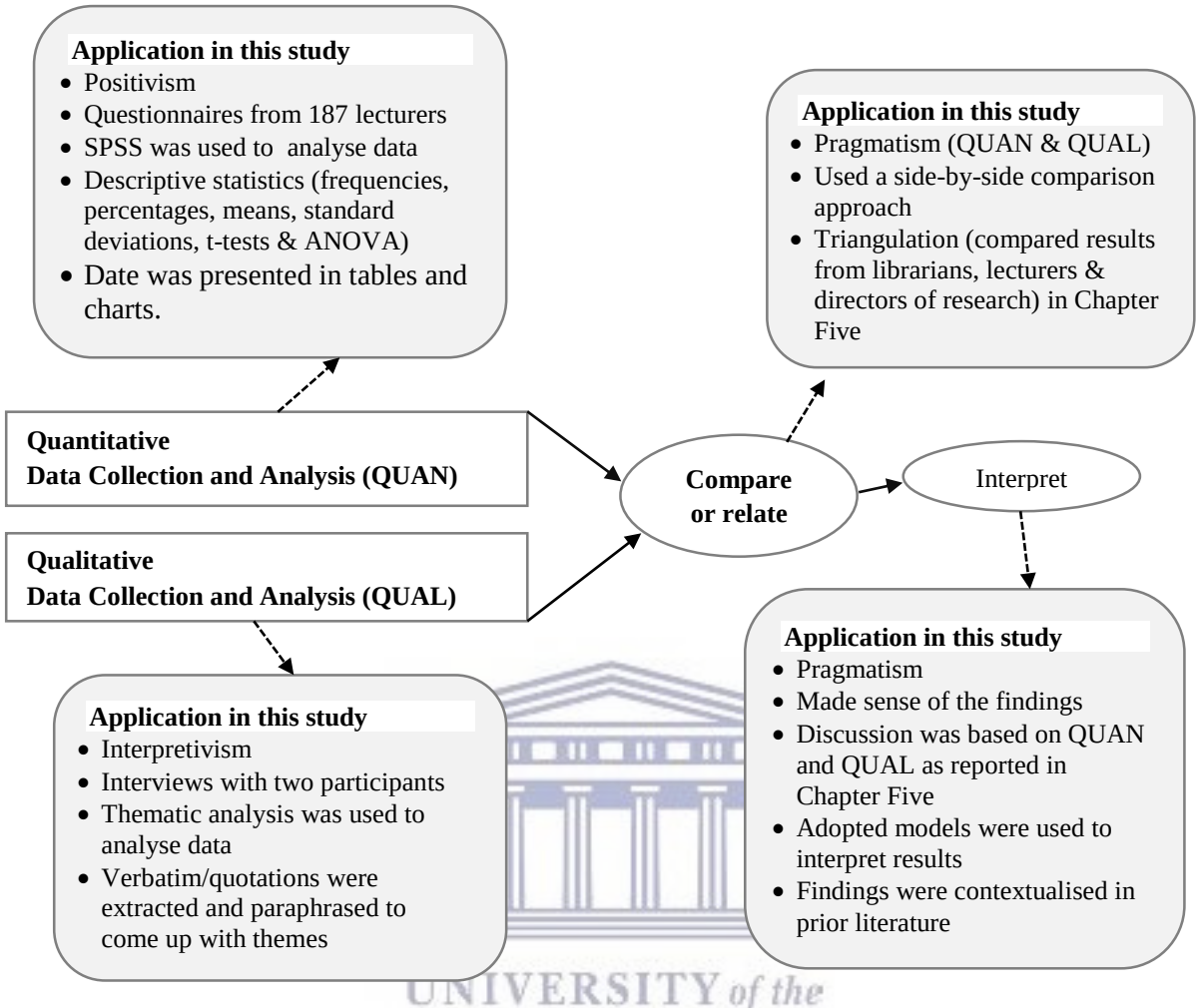


Figure 5. The convergent mixed methods design. Adapted from Creswell and Plano Clark (2011, p. 69) and Creswell (2014, p. 220).

Note.

→ = Depiction of convergent mixed methods design by Creswell and Plano Clark (2011, p. 69) and Creswell (2014)

⇢ = Depiction of the convergent mixed methods design in this research.

QUAN = Quantitative; QUAL = Qualitative.

4.7. Positioning methodology of the current study in relation to previous studies

Most research decisions interwoven into this study are informed by related prior studies. Particular aspects include the research paradigm, study population, data collection instruments and data collection procedures, just to mention a few. For instance, on the data collection instrument (questionnaire), most research variables were adopted from various authors such as Tenopir et al. (2011). These studies and their methodological approaches are summarised in Table 8. Noting that "philosophical backgrounds [paradigms] usually remain implicit in most

research" (Wahyuni, 2012, p. 69), the researcher had to read between the lines to identify the type of the paradigm(s) the studies employed. For instance, where the researchers used questionnaires with closed-ended questions and interviews to collect data, it was implied that a pragmatic paradigm was adopted.

Table 8. Methodology of related studies

Authors and study focus	Country	Study population	Data collection instruments and research paradigms
Digital curation in the Italian context (Brambilla, 2015).	Italy	Librarians	Interviews with five librarians in Italy (Interpretivism)
Survey on the needs for chemistry research data management and sharing (Chen & Wu, 2017).	China	Chemistry researchers	Questionnaires were sent to 119 researchers at the Chinese Academy of Science (Positivism)
Research data management and libraries (Cox & Pinfield, 2016).	UK	Librarians	Questionnaires were sent to 116 librarians in research institutions (Positivism)
The user's view on biodiversity data sharing (Enke et al., 2012)	Germany, USA, Austria, Canada	Biodiversity science researchers	Interviews with over 60 researchers and questionnaires were sent to 700 researchers. (Pragmatism)
Research data management in South Africa (Kahn et al., 2014).	South Africa	Librarians	Questionnaires (Positivism)
Data archiving, management initiatives and expertise in the Biological Sciences Department (Koopman, 2015).	South Africa	Biodiversity researchers	Face-to-face interviews with retired researchers and technical staff and questionnaires sent to researchers (Pragmatism)
An evaluation of a structured training event aimed at enhancing the research data management knowledge and skills of library professionals in South African higher education institutions (HEIs) (Matlatse, 2016).	South Africa	University librarians	Questionnaire was sent to workshop attendees of RDM (Positivism)
Research Data Management: An exploration of the data librarian role in New Zealand research organisations (Ohaji, 2016).	Australia	Research librarians	Interviews were conducted with research librarians (Interpretivism)
Data curation perspectives and practices of researchers at Kent State University's Liquid Crystal Institute: A case study (Shakeri, 2013).	USA	Researchers	Interviews and questionnaires (Pragmatism)

4.8. Data storage and archiving

Quantitative data captured in SPSS and qualitative data transcribed in text were archived in the University of the Western Cape repositories where it was made freely accessible to the general public so that it can be accessed and re-used by other researchers. The data is available at https://uwc.figshare.com/s/a040a0296a4337d6f32e. All details identifying the research participants were removed during the data cleaning process before the data was archived and released to the public for re-use.

4.9. Ethical issues

Gray (2009) defines research ethics as the appropriateness of the researcher's behaviour in relation to the subjects of the research or those who are affected by it. Ethics entail approval to conduct a study, risks, rights and dignity, and confidentiality of participants (Walliman, 2006). The study commenced only after being scrutinised and cleared by the University of the Western Cape's Humanities and Social Sciences Research Ethics Committee. As required by the University of the Western Cape's Humanities and Social Sciences Research Ethics Committee, informed consent was sought in writing from participants before data collection started. The purpose of the study was fully explained to the participants in writing when seeking consent. During the actual data collection, the researcher did not demand names from the participants and instead, numbers were used for the questionnaires. The researcher is convinced that most lecturers and librarians accepted participation in the study because they were aware that their responses and quotes could be used without their identity being known.

There are two key ethical issues to be considered when dealing with interviews, that is, anonymity (confidentiality) and sensitivity (Allmark et al., 2009; Cater, 2014; De Vos, Strydom, Fouche, & Delport, 2005). In terms of anonymity, this researcher adhered to Cater's (2014) recommendation that if the researcher wants to record an interview, it is highly recommended to seek permission from the participants prior to the commencement of the interview. Even if participants remain anonymous by removing all information that may identify them, use of quotes in writing up reports may reveal the identity of some participants (Allmark et al., 2009). In terms of sensitivity, the researcher followed advice from De Vos et al. (2005, p. 289) who warn researchers to refrain from asking participants questions which are personal in nature. Since Davies (2006) recommends that, whilst the participants may agree to participate in a study, they should nevertheless also feel free and be free to exercise their powers of veto during the research process. The researcher granted the participants full rights to

withdraw from the study at any stage of the research process if they wished to do so without giving reasons. See Appendices D, E and F for details about information consent letters. Fortunately, both participants agreed to participate in the study by signing the consent form and the interviews were recorded. The research sites were both anonymised using pseudonyms. The University offering general disciplines was named university 1, abbreviated as UNI1 while the university offering health related disciplines was named Universality 2 abbreviated as UNI2.

4.10. Summary of the chapter

Chapter Four has chronicled the methodology in this research. The study discussed research paradigms and it settled for the pragmatic paradigm. Because the pragmatic paradigm is the basis for mixed methods research, the study embraced a mixed methods research deign and more precisely, the convergent mixed methods deign was adopted. Based on a mixed methods design, the study collected both quantitative and qualitative data which was used to validate each other. Purposive sampling was used to select research sites and directors of research with whom the researcher conducted interviews. A census which is a non-sampling technique was used to identify lecturers and librarians. Quantitative data was collected from lecturers and librarians using a questionnaire whereas an interview guide was used to conduct in-depth interviews with directors of research. Quantitative research was analysed using SPSS whereas thematic analysis was applied on analysing qualitative data. The chapter also discussed issues of research reliability and validity; and ethical issues. Chapter Four (Data Presentation and Analysis) which is next, will present the findings of the empirical study.

CHAPTER FIVE

DATA PRESENTATION AND ANALYSIS

5.1. Introduction

This chapter presents findings from data that was collected using two data collection instruments, namely, a questionnaire and interviews. Creswell (2008) defines data presentation and analysis as a series of steps that researchers perform with the aim to identify key themes from the study that culminate in a discussion of the findings. Lunenburg and Irby (2008, p. 206) advise that when writing a presentation and analysis chapter of a dissertation, postgraduate students should avoid making references to the literature and the theoretical framework as doing so will detract from the purpose of the chapter. In this spirit, the researcher focusses on presenting the findings of the study without necessary relating them to either the prior literature or the models that underpin the study.

The aim of the study was to investigate how research data is generated, shared, stored and preserved in medical sciences and humanities for the purpose of re-use and long-term access. Data was collected from two public universities and from three categories of respondents that included library staff, researchers and directors of research. While a questionnaire was used to collect data from library staff and researchers, interviews were conducted with directors of research from each university. Data has been presented following research themes that were devised in Chapter One. Four research objectives were formulated to help answer the research problem as follows:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers needed to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affected the management of research data in public universities in Malawi.

The chapter is divided into three parts: the first section presents quantitative data collected from library staff followed by quantitative data collected from researchers; the final section presents and analyses qualitative data collected from the directors of research.

5.2. Analysis of quantitative data from library staff

Eighteen questionnaires were sent to 18 library staff at UNI2 and 16 (88.8%) questionnaires were returned. Another set of 22 questionnaires was sent to library staff at UNI1 and 20 (90%) were retuned. Data presented in this section focusses on background information of respondents, research data creation practices, preservation practices of research data, competency in data curation activities and factors that affect research data management.

5.2.1. Background information

This section reports on various personal information of respondents that include institutional affiliation, gender, rank and qualification.

5.2.1.1. Respondents by institution

The study attempted to establish institutions to which responded were affiliated. Findings revealed that there were 16 (44.4%) respondents from UNI2 and 20 (55%) respondents from UNI2. These results show that there were more respondents from UNI1 than UNI2. This is because UNI1 is a fully fledged and a standalone university whereas UNI2 is a college under the University of Malawi.

5.2.1.2. Respondents by gender, qualification and rank

Respondents were asked to indicate their gender, qualification and rank. For each variable, means, standard deviations and t-tests were computed to determine if there were statistically significant difference between the two universities. Findings are presented in Table 9.

Table 9. Gender, qualification and rank (N = 36)

Dimension	Institution		*f(%)*	Mean	SD	t-value	Sig.
Gender	UNI2	Male	10(62.5%)	1.5625	.81394	.859	.396**
		Female	3(18.8%)				
		No response	3(18.8%)				
	UNI1	Male	15(75.0%)	1.3500	.67082		
		Female	3(15.0%)				
		No response	2(10.0%)				
Qualification	UNI2	Diploma	4(25.0%)	2.6875	1.44770	-.026	.979**
		Certificate	4(25.0%)				
		Bachelor's	3(18.8%)				
		Masters	2(12.5%)				
		PhD	0(0%)				
		No response	3(18.7%)	2.7000	1.38031		
	UNI1	Diploma	4(20.0%)				
		Certificate	6(30.0%)				
		Bachelor's	5(25.0%)				
		Masters	3(15.0%)				
		PhD	1(2.8%)				
		No response	2(5.6%)				
Rank	UNI2	Senior Library Assistant	7(43.8%)	2.1875	1.04682	.260	.796**
		Library Assistant	4(25.0%)				
		Assistant Librarian	4(25.0%)				
		Senior Assistant Librarian	0(0%)				
		University/College Librarian	1(6.2%)				
	UNI1	Senior Library Assistant	8(40.0%)	2.1000	.96791		
		Library Assistant	6(30.0%)				
		Assistant Librarian	4(20.0%)				
		Senior Assistant Librarian	2(10.0%)				
		University/College Librarian	0(0%)				

Note. **($p > 0.5$), constant interval percentage = 95%.

Results presented in Table 9 show that there were 10 (62.5%) males at UNI2 and three (18.8%) females and three (18.8%) did not indicate their gender. At UNI1, there were 15 (75%) males and three (15%) females and two (10%) did not indicate their gender. Analysis findings presented in Table 9 reveals that there were more male than female respondents. An independence t-test was computed to determine if there were statistically significant differences in gender between the two universities. Means and standard deviations revealed that there were no significant differences with a lower mean (M =1.3500) and the higher mean (M = 1.5625) and the lower standard deviation was (SD = .67082) and the higher standard deviation was (SD = .81394). Analysis of the p-value confirmed further that there were no statistically significant differences ($p > 0.5$) in gender between the two universities (t (34) = .859, p = .396).

Pertaining to respondents' qualification, Table 9 reveals that at UNI2, four (25%) had a diploma, another four (25%) had a certificate, three (18.8%) had bachelor's degree, two (12.5%) had a master's degree and three (18.8%) did not respond. At UNI1, four (20%) had a diploma, six (30%) had a certificate, five (25.0%) had a bachelor's degree, three (15%) had a master's degree, one (2.8%) had a doctorate degree and two (5.6%) did not respond. An analysis of means (M = 2.6875 vs. 2.7000) and standard deviations (SD = 1.44770 vs. 1.38031) show that there were no significant differences in qualifications between the two universities and standard deviation. Likewise analysis of the p-value showed there were no statistically significant differences ($p > 0.5$) between the two universities ($t(340 = -.026$, $p = .979$).

In terms of rank, results in Table 9 show that at UNI2, there were seven (43.8%) senior library assistants, four (25%), four (25%), assistant librarians and one (6.2%) college librarian while at UNI1, there were eight (40%) senior library assistants, six (30.0%) library assistants and two (10.0%) senior assistant librarians. Highest and lowest means and standard deviations in that order resulting from the computation of the t-test were (M = 2.1875, 2.1000; SD = 1.04682, .96791) respectively meaning there were no significant differences in terms of rank at the two universes. Analysis of the p-value showed further that there were no statistically significant differences ($p = > 0.5$) in rank between the two universities ($t(34) = .260$, $p = .796$).

5.2.2. Research data creation practices

This section addressed a number of questions with the aim to understand the research data creation, sharing and re-use practices in public universities in Malawi.

5.2.2.1. Researches consultation with library staff on research activities

In the first item of the questionnaire for library staff (See Appendix A), the researcher wanted to know if researchers consulted them on research issues. Findings are presented in Figure 6. Findings reveal that the majority of library staff with scores of 27 (75%) indicated that researchers consulted them on research activities. Only seven (19.4%) said they were not consulted while two (5.6%) did not respond to this question. The results generally suggest that researchers consult library staff in research activities.

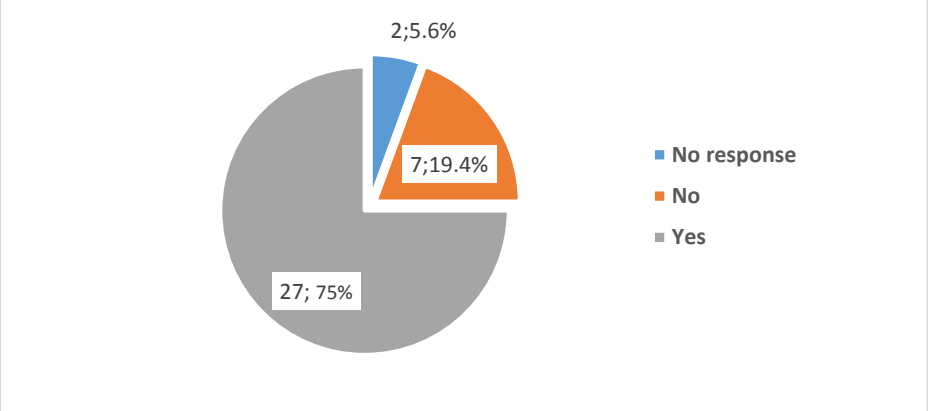


Figure 6. Researcher's consultation with library staff on research (N = 36)

5.2.2.2. Research areas on which researcher consult library staff

A follow up question required librarians to mention the research activities on which researchers approached them for help. The findings are presented in Table 10.

Table 10. Research areas on which researchers consult library staff (N = 36)

Research activities	Frequency	Percentage
Citation and referencing	30	83.3
Identification of creditable journals	26	72.2
Installation of data analysis software (e.g. SPSS)	26	72.2
Data collection	24	66.7
Data storage and preservation	18	50.0
Data analysis using computer software	14	38.9
Sources of funding opportunities	6	16.7
Developing online data collection tools	6	16.7
Research areas	6	16.7
Recovery of lost research data/information	3	8.3
Data cleaning	0	0
Sources of research collaboration	0	0

Results presented in Table 10 show that 30 (83.3%) helped researchers in citation and referencing; 26 (72.2%) in identification of creditable journals; 26 (72.2%) in installation of data analysis software; 24 (66.7%) in data collection; 18 (50%) in data storage and preservation; 14 (38.9%) in data analysis using computer software; six (16.7%) in developing online data collection tools, six (16.7%) in research areas and three (8.3%) in recovery of lost research data/information. Findings show further that researchers did not consult library staff in data cleaning and sources of research collaboration.

The results suggest that generally, researchers consult library staff on various research activities but the most predominant ones include citation and referencing; identification of creditable journals; installation of data analysis software; data collection; and data storage and preservation.

5.2.3. Preservation practices

This section of the library staff questionnaire aimed at investigating preservation practices and availability of research data infrastructure.

5.2.3.1. Digital storage facilities available in libraries for use by researchers

One item in this section requested library staff to indicate technological facilities that were available in libraries for use by researchers to manage research data. Means, standard deviations and t-tests were computed to explore if there were significant differences in storage facilities between the two universities. Findings are presented in Table 11.

Table 11. Technological/digital storage facilities (N = 36)

Storage facility	University	N	*f* & (%)***	M	SD	t-value	Sig.
Office computers	UNI2	16	16 (100%)	1.0000	.00000	-1.944	.060**
	UNI1	20	16 (80%)	1.2000	.41039		
	Total	36	32(88.9%)	-	-		
External hard drives	UNI2	16	12(75.0%)	1.2500	.44721	-1.850	.073**
	UNI1	20	9 (45.0%)	1.5500	.51042		
	Total	36	21(58.3%)	-	-		
CDs	UNI2	16	0 (0%)	2.0000	.00000	**a**	**a**
	UNI1	20	0 (0%)	2.0000	.00000		
	Total	36	0 (%)	-	-		
Institution's available networked capacity	UNI2	16	15 (93.8%)	1.0625	.25000	-2.776	.009*
	UNI1	20	11(55.0%)	1.4500	.51042		
	Total	36	26(72.2%)	-	-		
Commercial software or services	UNI2	16	0 (0%)	2.0000	.00000	**a**	**a**
	UNI1	20	0 (0%)	2.0000	.00000		
	Total	36	0 (%)	-	-		
Freely available software or services (Google Drive)	UNI2	16	15(93.8%)	1.0625	.25000	-4.360	.000*
	UNI1	20	7(35.0%)	1.6500	.48936		
	Total	36	22(61.1%)	-	-		
Flash disks	UNI2	16	15(93.8%)	1.0625	.25000	-.394	.696**
	UNI1	20	18(90.0%)	1.1000	.30779		
	Total	36	33 (91.7%)	-	-		
Email account(s)	UNI2	16	11(68.8%)	1.3125	.47871	2.179	.036*
	UNI1	20	19(95.0%)	1.0500	.22361		
	Total	36	30(83.3%)	-	-		

Note. . [a.] t was not computed because the standard deviations of both groups were 0, * ($p \leq 0.5$), ** ($p > 0.5$), ***percentages calculated against total number of individual universities i.e. % of 20 for UNI1 and % of 16 for UNI2, constant interval percentage = 95%.

Results in Table 11 show that all 16 (100%) UNI2 and 16 (80%) UNI1 library staff indicated that office computers were available; 12 (75%) UNI2 and nine (45%) mentioned external hard drives, 15 (93.8%) UNI2 and 11 (55%) UNI1 staff mentioned institution's networked capacity; 15(93.8%) UNI2 and seven (35%) UNI1 staff indicated that there were freely available software or services such as Google Drive; 15 (93.8%) UNI2 and 18 (90%) UNI1 mentioned flash disks; and 11 (68.8%) UNI2 and 19 (95%) UNI1 staff indicated that email accounts were provided. These results suggest that flash discs, computers, email accounts, networks, free software and external hard drives were the common technological storage facilities available

in libraries of both universities. Although results show that libraries had various digital storage facilities that could be used by researchers to store and preserve research data, these libraries did not have commercial software or services.

Independent t-test results revealed that while there were statistically significant differences in some storage facilities, there were no statistically significant differences in others. Specifically, there were statistically significant differences in institution's available networked capacity (M = 1.0625 vs. 1.4500; SD = .25000 vs. .51042; $t(34) = -2.776$, $p = .009$); freely available software or services (M = 1.0625 vs. 1.6500, SD = 1.0625 vs. 1.6500; $t(34) = -4.360$, $p < .001$); and email accounts (M = 1.3125 vs. 1.0500, SD = .47871 vs. .22361; $t(34) = 2.179$, $p = .036$). On the other hand, there were no significant differences ($p > 0.5$) in office computers (M = 1.0000 vs. 1.2000, SD = .00000 vs. .41039; $t(34) = -1.944$, $p = .060$); external hard drives (M = 1.2500 vs. 1.5500, SD = .44721 vs. .51042; $t(34) = -1.850$, $p = .073$); and flash disks (M = 1.0625 vs.1.1000, SD = .25000 vs. .30779; $t(34) = -.394$, $p = .696$).

5.2.3.2. Helping researchers to back up research data

Library staff were asked to indicate their capability in helping researchers to make backups or preserving research data. Responses are presented in Table 12.

Table 12. Library staff's ability to help researchers make backups of their research data (N = 36)

Type of help	Frequency	Percentage
Helping to save copies on a local server		
I do help them already	1	2.8
I have not helped them before but I am ready to help	10	27.8
I am not ready to help because I lack skills	16	44.4
Not sure	9	25
Helping to save files on a disk, USB drive, computer hard drive		
I do help them already	16	44.4
I have not helped them before but I am ready to help	19	52.8
I am not ready to help because I lack skills	1	2.8
Not sure	0	0
Helping to save data files on a central campus server		
I do help them already	0	0
I have not helped them before but I am ready to help	8	22.2
I am not ready to help because I lack skills	5	13.9
Not sure	1	2.8
No response	22	61.1
Helping to save data files on a web-based or cloud server		
I do help them already	5	13.9
I have not helped them before but I am ready to help	20	55.6
I am not ready to help because I lack skills	0	0
Not sure	10	27.8
No response	1	2.8
Helping to store copies in repositories		
I do help them already	22	61.1
I have not helped them before but I am ready to help	8	22.2
I am not ready to help because I lack skills	5	13.9
Not sure	0	0
No response	1	2.8
Restricting access to files		
I do help them already	0	0
I have not helped them before but I am ready to help	9	25.0
I am not ready to help because I lack skills	5	13.9
Not sure	20	55.6
No response	2	5.6

Analysis of data presented in Table 12 shows that the majority of library staff with scores of 16 (44.4%) lacked skills in helping researchers to save research files on a local server, 10 (27.8%) were ready to help researchers but had not done so before but they had the capability. Only one (2.85) library had helped researchers in saving copies on local server whereas nine (25%) were not sure. In terms of helping save files on a disk, USB drive, computer hard drive, 19 (52.8%), had not helped researcher before, 16 (44.4%) were already helping researchers,

and one (2.8%) had not helped researchers before. None of the library staff indicated had helped researchers in saving files on a central campus server, eight (22.2%) said they were ready but they had not helped before, five (13.9%) said they lacked skills while 22 (61.1%) said they were not sure. Pertaining to helping researchers save data files on a web-based or cloud server, five (13.9%) said they were already helping researchers, 20 (55.6%) said they were ready but had not helped before, none said lacked skills and 10 (27.8%) said were not sure. Five (13.9%) library staff had helped researchers to save files on web-based cloud servers, 20 (55.6%) were ready to help them, none of the library staff lacked skills and 10 said they were not sure. Twenty two (61.1%) library staff were already helping researchers to store copies in a repository, eight (2.2%) were ready to help and five (13.9%) lacked skills. In terms of restricting access to files, none of the library staff helped researchers, nine (25%) said they were ready to help, five (13.9%) said they were not ready to help because they lacked skills and 20 (55.6%) said were not sure.

An analysis of the results suggests the majority of the library staff were already helping researchers in storing copies in a repository or archives and the other majority was ready to help researchers in saving files on a disk, flash disks, computer hard drives and saving data files on a web-based or cloud server. Worth noting is that most library staff did not help researchers in saving copies on a local server because they lacked skills.

5.2.3.3. Decisions regarding research data preservation

In this question, library staff were asked to indicate the extent to which they helped researchers in making decisions about data preservation. Findings are depicted in Table 13.

Table 13. Decisions regarding research data preservation (N = 36)

Services	Frequency	Percentage
Deciding which data is important to preserve		
I do help them already	3	8.3
I have not helped them before but I am ready to help	0	0
I am not ready to help because I lack skills	26	72.2
Not sure	7	19.4
Deciding whether data can be safely shared		
I do help them already	2	5.6
I have not helped them before but I am ready to help	4	11.1
I am not ready to help because I lack skills	30	83.4
Not sure	0	0
Determining standards for identifying sensitive data		
I do help them already	0	0
I have not helped them before but I am ready to help	1	2.8
I am not ready to help because I lack skills	21	58.3
Not sure	14	38.9
Determining what constitutes compliance with commercial licenses, government regulations, funding agency mandates		
I do help them already	3	8.3
I have not helped them before but I am ready to help	8	22.2
I am not ready to help because I lack skills	17	47.2
Not sure	8	22.2
Determining appropriate metadata to describe data sets (i.e., descriptive information to enable others to reuse data)		
I do help them already	3	8.3
I have not helped them before but I am ready to help	8	22.2
I am not ready to help because I lack skills	17	47.2
Not sure	8	22.2
Determining provisions for short and long-term data storage/preservation		
I do help them already	1	2.8
I have not helped them before but I am ready to help	9	25
I am not ready to help because I lack skills	22	61.1
Not sure	4	11.1

Data presented in Table 13 shows that only three (8.3%) library staff already helped researchers when deciding which data is important to preserve, none of the library staff were ready to help and 26 (72.2%) said they lacked skills in this aspect and seven (19.4%) said they were not sure. In terms of deciding which data could be safely shared, only two (5.6%) were already helping researchers, 30 (83.4%) said they lacked skills and four (11.4%) said they were not sure. None of the library staff were already helping researchers in determining standards for identifying sensitive data, one (2.8%) librarian was ready to help researchers in this aspect and 21 (58.3%)

said they lacked skills. In terms of helping researchers comply with licenses, regulations and mandates on data management, only three (3.3%) library staff indicated they were already helping researchers, eight (22.2%) said they were ready to help, 17 (47%) said they did not have skills and eight (22.2%) said they were not sure. Three (8.3%) library staff indicated that they already helped researchers to assign metadata, eight (22.2%) said they were ready to help, 17 (47%) said they lacked skills and eight (22.2%) said were not sure. Only one (2.8 %) library staff member was already helping researchers in determining data storage and preservation on long term, nine (25%) said they were ready to help, 22 (61.1%) said they lacked skills in this aspect and four (11.15%) said they were not sure.

Generally, analysis of the results suggest that the majority of library staff lacked skills in all aspects investigated in this section, that is, helping researchers decide which data is important to preserve; deciding which data can be safely shared; determining standards for identifying sensitive data; helping comply with licenses, regulations and mandates on data management; assigning metadata; and determining data storage and long-term preservation.

5.2.3.4. Provision of data preservation services to researchers

In this item, library staff were asked to express their opinions about their libraries' readiness in the provision of research data management services. Findings are presented in Table 14.

Table 14. Library's readiness in the provision of data preservation services to researchers (N = 36)

Service	Frequency	Percentage
Provision of advanced computing options		
The library is already offering	25	69.4
The library is not offering but it has capability	3	8.3
The library does not have the capability to offer	0	0
Not sure	7	19.4
No response	1	2.8
Provision of statistical and other data analysis support		
The library is already offering	11	30.6
The library is not offering but it has capability	4	11.1
The library does not have the capability to offer	11	30.6
Not sure	10	27.8
Short and long-term data storage/preservation		
The library is already offering	2	5.6
The library is not offering but it has capability	20	55.6
The library does not have the capability to offer	5	13.9
Not sure	8	22.2
No response	1	2.8
Data security support		
The library is already offering	4	11.1
The library is not offering but it has the capability	7	19.4
The library does not have the capability to offer	2	5.6
Not sure	18	50.0
No response	5	13.9
Guidance on how to use appropriate metadata		
The library is already offering	9	25.0
The library is not offering but it has the capability	7	19.4
The library does not have the capability to offer	1	2.8
Not sure	19	52.8
Guidance on writing a data management plan		
The library is already offering	5	13.9
The library is not offering but it has the capability	3	8.3
The library does not have the capability to offer	2	5.6
Not sure	26	72.2

Findings presented in Table 14 reveal that 25 (69.4%) library staff indicated that libraries were already offering advanced computing options, three (8.3%) said libraries have the capability, none of the respondents said libraries lacked the capability and seven (19.4%) said were not sure. One (2.8%) respondent did not answer this question. On the aspect of statistical and other

data analysis support, 11(30.6%) respondents said libraries were already offering such support, four (11.1%) said libraries had the capability, 11 (30.6%) said libraries did not have capability and 10 (27.8%) said were not sure. Two (5.6%) respondents were of the view that libraries were already offering short and long-term data storage/preservation facilities to researchers, 20 (55.6%) indicated that libraries had the capability to offer this service, five (13.9%) were of the view that libraries lack the capability, eight 8 (22.2%) were not sure and one (2.8%) library staff did not respond to this question. For the service of data security support, four (11.1%) respondents indicated that libraries were already offering, seven (19.4%) said libraries had the capability to offer this service, two (5.6%) said libraries did not have the capability, 18 (50%) were not sure and five (13.9%) did not comment. Nine (25%) respondents indicated that libraries were already helping researchers in the use of appropriate metadata, seven (19.4%) were of the view that libraries had the capability, one (2.8%) indicated that libraries did not have a capability and 19 (52.8%) said they were not sure. Finally, on libraries' readiness to help researchers in preparing data management plans, three (8.3%) said libraries were already helping researchers, and 26 (72.2%) said libraries had the capability to offer this service, two (5.6%) were of the view that libraries did not have a capability and five (13.9%) indicated they were not sure.

Analysis of the findings reveals that libraries were already offering advanced computing options and statistical and other data analysis support. Though not actively offering, libraries had capabilities in offering short and long-term data storage/preservation facilities to researchers and preparing data management plans. On the other hand, the majority of library staff were not certain if libraries were ready to provide data security support and the use of appropriate metadata.

5.2.3.5. Research data infrastructure

Library staff were asked to express their opinion if the universities had enough infrastructure to support management of research generated within the university. Findings are presented in Figure 7 where it is clear that an equal number of respondents (18 or 50%) agreed to the question as (18 or 50%) disagreed.

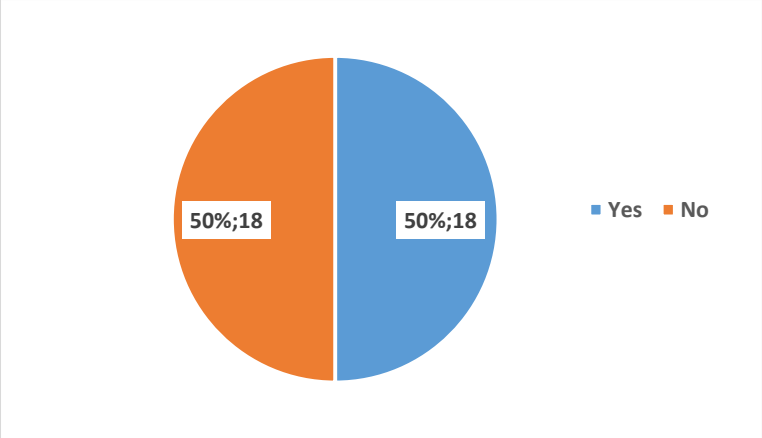


Figure 7. Library staff's opinion about research data management infrastructure (N = 36)

5.2.3.6. Support for research data management

A follow-up question to the preceding one required library staff to indicate the type of support that their universities needed to provide in order to strengthen research data management activities. This question was exclusive to respondents who answered no to the preceding one; however, the researcher noted that the majority of those respondents who answered yes provided answers to this question. Findings are presented in Table 15. ANOVA was also computed to explore differences in the kind of support by university affiliation.

Table 15. Kind of support to strengthen research data management (N = 36)

Support	Institution	Agree strongly	Agree somewhat	Neutral	Disagree somewhat	Disagree strongly	No response
Establish process for managing data for 5 years or less[1]	UNI2	4 (25%)	0(0%)	0(0%)	0(0%)	12 (75%)	0(0%)
	UNI1	9 (45%)	6 (30%)	0(0%)	0(0%)	0(0%)	5(25.5%)
Establish a process for managing data beyond 5 years[2]	UNI2	4 (25%)	0(0%)	0(0%)	0(0%)	12 (75%)	0(0%)
	UNI1	10(50%)	4 (20%)		0(0%)	1(1%)	5 (25.5%)
Establish technical support for data management[3]	UNI2	12(75%	0(0%)	0(0%)	0(0%)	4 (25%)	0(0%)
	UNI1	6 (30%)	8 (40%)			1(1%)	5(25.5%)
Establish funds to support data management[4]	UNI2	4 (25%)	0(0%)	0(0%)	0(0%)	12 (75%)	0(0%)
	UNI1	7 (35%)	7 (35%)	0(0%)		1(1%)	5(25.5%)

ANOVA

[1]($F(1, 34) = 9.261$, $p = .004$); [2]($F(1, 34) = 8.051$, $p = .008$); [3]($F(1, 34) = 7.014$, $p = .012$); [4]($F(1, 34) = 7.264$, $p = .011$).

Note. Constant interval percentage = 95%

Findings presented in Table 15 show that at UNI2, four (25%) library staff agreed strongly, and 12 (75%) disagreed strongly on the need to establish a process for managing data for five years or less. On the other hand, on the same aspect, 12 (45%) and nine (30%) UNI1 respondents agreed strongly and agreed somewhat respectively while five (25%) did not comment. On the need to establish a process for managing data beyond five years, four (25%) UNI2 respondents strongly agreed while 12 (75%) disagreed strongly while 10 (50%) UNI1 agreed strongly, four (20%) agreed somewhat, one (1%) disagreed strongly and five (25.5%) did not comment. Twelve (75%) and four (25%) UNI2 respondents agreed strongly and disagreed strongly respectively on the on the need to establish technical support for data management while for UNI1 staff, six (30%) agreed strongly, eight (40%) agreed somewhat, one (1%) disagreed strongly and five (25.5) did not comment. On the aspect of establishing funds to support data management, four (25%) UNI2 staff strongly agreed and 12 (75%) disagreed strongly and for UNI1 respondents, seven (35%) agreed strongly, another seven (35%) agreed somewhat, one (1%) disagreed strongly and five (25.5%) did not comment. Analysis of the findings shows that while UNI2 library staff were of the opinion that their university should focus on establishing technical support for data management, UNI1 library staff were of the view that their university should introduce all these kinds of support.

ANOVA results show that there were statistically significant differences in all dimensions between the two universities namely; establish process for managing data for five years or less ($F(1, 34) = 9.261$, $p = .004$); establish a process for managing data beyond five years ($F(1, 34) = 8.051$, $p = .008$); establish a process for managing data beyond 5 years ($F(1, 34) = 7.014$, $p = .012$); and establish funds to support research data management ($F(1, 34), = 7.264$, $p = .011$).

5.2.4. Competency in data curation activities

This section aimed at investigating if library staff possessed appropriate skills for managing data.

5.2.4.1. Training workshops on research data management

This item required library staff to indicate if they had ever attended workshops in research data management. Findings are presented in Figure 8 where it is clear that only 10 (27.8%) respondents had attended workshops while 26 (72.2%) said they had not. A cross-tabulation of the results revealed that of the 10 (27.8%) who attended the workshops, three (30%) were from UNI1 and seven (70%) were from UNI2.

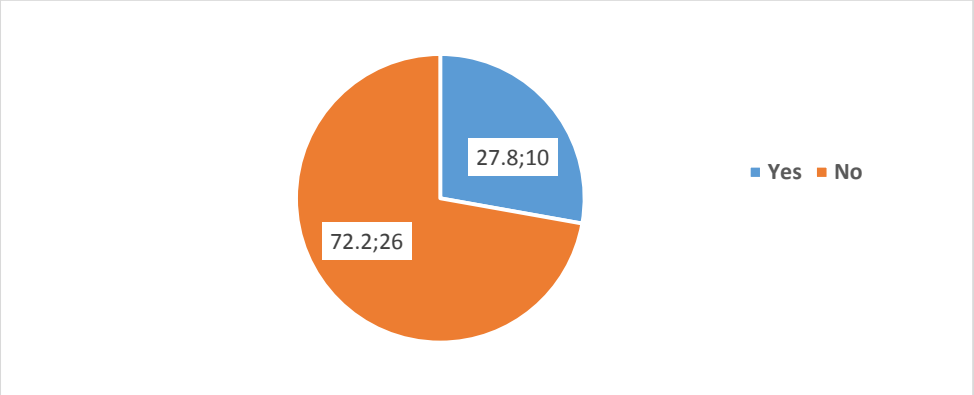


Figure 8. Training workshops on research data management (N =36)

A follow-up question to the preceding one required respondents to indicate the organisers of the workshops. Six (16.7%) said they were organised by their universities while four (11.1%) said they were organised by international organisations. This means that local agencies did not organise such workshops for librarians.

5.2.4.2. Specific competencies in managing research data

Respondents, provided with a list of activities in managing research data, were asked to indicate whether they were competent or if they needed to be trained by experts. Findings are presented in Table 16.

Table 16. Specific competencies in managing research data (N = 36)

Dimension	University	N	Competent***	Need training***	M	SD	t-value	Sig.
Identifying new standards and practices for curation	UNI2	16	7 (43.8%)	9(56.2%)	1.5000	.51640	-1.214	.233**
	UNI1	20	6(30%)	14(70.0%)	1.7000	.47016		
	Total	36	13 (36.9%)	23 (63.1%)	-	-		
Curating digital objects using curation lifecycle	UNI2	16	6(37.5%)	10(62.5%)	1.5000	.51640	-.892	.379**
	UNI1	20	7(35.0%)	13(65%)	1.6500	.48936		
	Total	36	13 (36.9%)	23 (63.1%)	-	-		
Long term digital data preservation strategies	UNI2	16	7(43.8%)	9 (56.2%)	1.5625	.51235	-.838	.408**
	UNI1	20	6(30%)	14(70%)	1.7000	.47016		
	Total	36	13(36.1%)	23 (63.9%)	-	-		
Creating preservation metadata	UNI2	16	9(56.2%)	7 (43.8%)	1.4375	.51235	-.363	.719**
	UNI1	20	10(50%)	10(50.0%)	1.5000	.51299		
	Total	36	19(52.8%)	17(47.2%)	-	-		
Collecting data from creators	UNI2	16	12(75%)	4(25%)	1.2500	.44721	-2.179	.036*
	UNI1	20	12(60%)	8 (40.0%)	1.6000	.50262		
	Total	36	24(66.7%)	12 (33.3%)	-	-		
Selecting digital objects for preservation	UNI2	16	7(43.8%)	9(56.2%)	1.5625	.51235	-.221	.827**
	UNI1	20	8(40%)	12(60%)	1.6000	.50262		
	Total	36	15(41.7%)	21(58.3%)	-	-		
Transferring preserved digital objects to repositories	UNI2	16	10(62.5%)	6(37.5%)	1.3750	.50000	-1.338	.190**
	UNI1	20	8(40%)	12 (60%)	1.6000	.50262		
	Total	36	18(50%)	18(50%)	-	-		
Storing digital information using standards	UNI2	16	6(37.5%)	10(62.5%)	1.3750	.50000	-1.659	.106**
	UNI1	20	7(35%)	13(65%)	1.6500	.48936		
	Total	36	13 (36.9%)	23 (63.1%)	-	-		
Providing access to stored digital objects to users	UNI2	16	12(75%)	4(25%)	1.2500	.44721	-2.179	.036*
	UNI1	20	12(60%)	12(40%)	1.6000	.50262		
	Total	36	20(55.6%)	16(44.4%)	-	-		
Disposing data not selected for long term preservation	UNI2	16	8(50%)	8(50%)	1.5000	.51640	-.892	.379**
	UNI1	20	7(35%)	13(65%)	1.6500	.48936		
	Total	36	15(41.7%)	21(58.3%)	-	-		
Migrating digital data to newer file formats	UNI2	16	9(56.2%)	7(43.8%)	1.5625	.51235	-.838	.408**
	UNI1	20	10 (50%)	10 (50%)	1.7000	.47016		
	Total	36	13(36.1%)	23(63.9%)	-	-		
Citing and transforming data	UNI2	16	5(31.2%)	11(68.8%)	1.3125	.47871	-1.426	.163**
	UNI1	20	9(45%)	11(55%)	1.5500	.51042		
	Total	36	14(38.9%)	22(61.1%)	-	-		

Note. * ($p \leq 0.5$), ** ($p > 0.5$), ***percentages calculated against total number of individual universities i.e. % of 20 for UNI1 and % of 16 for UNI2; percentages in the total row are calculated against the total number i.e. % of 36, constant interval percentage = 95%

According to the findings presented in Table 16, seven (43.8%) UNI2 staff were competent in identifying new standards and practices for curation and nine (56.2%) needed training while six (30%) UNI1 staff said they were competent and 14 (70%) said they needed training. On the aspect of curating digital objects using the curation lifecycle, six (37.5%) UNI2 staff said they were competent and 10 (62.5%) needed training while at UNI1, seven (35%) said were competent and 13 (65%) said they needed training. In terms of long term digital data

preservation strategies, seven (43.8%) UNI2 staff said they were competent while nine (56.2%) needed training and six (30%) UNI1 staff were competent while 14 (70%) said they needed training. Pertaining to creating preservation metadata, nine (56.2%) UNI2 staff said they were competent while seven (43.8%) said they needed training and for UNI1 staff, 10 (50%) were competent and another 10 (50%) said they needed training. On the skill of collecting data from creators, 12 (75%) UNI2 staff were competent and four (25%) needed training and at UNI1, 12 (60%) were competent while eight (40%) needed training. In terms of selecting digital objects for preservation, seven (43.8%) UNI2 staff said were competent and nine (56.2%) needed training while eight (40%) were competent and 12 (60%) said they needed training. On the aspect of transferring preserved digital objects to repositories, 10 (62.5%) said they were competent and six needed training while at UNI1, eight (40%) were competent and 12 (60%) said needed training.

Results in Table 16 show further that six (37.5%) UNI2 staff were competent in storing digital information using standards and 10 (62.5%) said they were not competent while seven (35%) UNI1 staff were competent and 13 (65%) were not competent. Twelve (75%) UNI2 staff were competent in providing access to stored digital objects to users and four (25%) said were not competent while 12 (60%) UNI1 staff were competent and eight (40%) were not competent. On the skill of disposing data not selected for long term preservation, eight (50%) UNI2 staff were competent and another eight (50%) were not competent while for UNI1 staff, seven (35%) were competent and 13 (65%) were not competent. In terms of migrating digital data to newer file formats, nine (56.2%) UNI2 staff said they were competent and seven (43.8%) were not competent and for UNI1, 10 were competent and another 10 (50%) were not competent. Finally, five (31.2%) UNI2 staff were competent in citing data and 11(68.8%) were not competent while for UNI1, nine (45%) were competent and 11 (55%) were not.

Analysis of the results show that more librarians at both universities were competent than incompetent in creating preservation metadata; collecting data from creators; providing access to stored digital objects to users; and migrating digital data to newer file formats. Results show further that library staff at both institutions need training in identifying new standards and practices for curation; curating digital objects using curation lifecycle; long term digital data preservation strategies; selecting digital objects for preservation; storing digital information using standards; citing and transforming data. Further analysis of results show that only UNI2 library staff were competent in transferring preserved digital objects to repositories and

disposing data not selected for long term preservation while UNI1 staff were not competent in these aspects.

Overall, results presented in Table 16 show that there were no statically significant differences in competencies by universities across all the 11 dimensions as evidenced by the scores of a minimum mean (M =1.2500) and maximum mean (M = 1.7000); and minimum standard deviation (SD = .44721) and maximum standard deviation (SD = .51640). Specifically, the following dimensions registered no statistically significant differences between the two universities: Identifying new standards and practices for curation ($t(34) = -1.214$, $p = .233$); curating digital objects using curation lifecycle ($t(34) = -.892$, $p = .379$); long term digital data preservation strategies ($t(34) = -.838$), $p = .408$); creating preservation metadata ($t(34) = -.363$, $p = .719$); selecting digital objects for preservation ($t(34) = -.221$, $p = .827$), transferring preserved digital objects to repositories ($t(34) = -1.338$, $p = .190$); storing digital information using standards ($t(34) = -1.659$, $p = .106$); disposing data not selected for long term preservation ($t(34) = -.892$, $p = .379$); migrating digital data to newer file formats ($t(34) = -.838$, $p = .408$); and citing and transforming data ($t(34) = 1.426$, $p = .163$). Results reveal that there were statistically significant differences ($p > 0.5$) in two dimensions, namely, collecting data from creators ($t(34) = -2.179$, $p = .036$) and providing access to stored digital objects to users ($t(34) = -2.179$, $p = .036$).

5.2.5. Challenges affecting research data management

The last item in the questionnaire required library staff to indicate the extent to which they agreed or disagreed with the factors that limited their involvement in research data management activities. Findings are presented in Table 17. ANOVA was also computed to explore differences in challenges between the two universities.

Table 17. Challenges affecting research data management

Factors	Institution of participant									
	UNI2					UNI1				
	Agree strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Failure by researchers to engage me in data curation [1]	13 (81.2%)	5 (12.5%)	0 (0%)	0 (0%)	1 (6.3%)	13 (65%)	4 (20%)	2 (10%)	1 (5%)	0(0%)
Lack of incentives to curate data [2]	1 (6.3%)	2 (12.5%)	3 (18.8%)	0(0%)	10 (62.5%)	1 (5%)	1 (5%)	11 (55%)	3 (15%)	4 (20%)
Larger amounts of data to curate [3]	1 (6.3%)	0(0%)	0(0%)	0(0%)	15 (93.8%)	0(0%)	0(0%)	0(0%)	11 (55%)	9 (45%)
Lack of time [4]	1(6.3%)	0 (0%)	0 (0%)	6 (37.5%)	9 (56.3%)	0 (0%)	0 (0%)	1 (5%)	8 (40%)	11 (55%)
Insufficient storage and network infrastructure [5]	1 (6.3%)	4 (25%)	4(25%)	7 (43.8%)	0 (0%)	12 (60%)	6 (30%)	2 (10%)	0(0%)	0(0%)
Lack of curation tools and software [6]	7 (43.8%)	5 (31.3%)	4 (25%)	0 (0%)	0 (0%)	5 (25%)	1 (5%)	12 (60%)	2 (10%)	0 (0%)
Lack of policy frameworks [7]	9 (56.3%)	7 (43.8%)	0 (0%)	0 (0%)	0 (0%)	9 (45%)	5 (25%)	2 (10%)	2 (10%)	2 (10%)
Lack of curation skills and training [8]	3 (18.8)	8 (50%)	0(0%)	2 (12.5%)	3 (18.8%)	15 (75%)	1 (5%)	0(0%)	4 (20%)	0(0%)
Lack of guidance and support [9]	1 (6.3%)	9 (56.3%)	3 (18.8%)	0 (0%)	3 (18.8%)	6 (30%)	7 (35%)	5 (25%)	2 (10%)	0 (0%)
Difficulty in accessing data[10]	11 (68.8%)	4 (25%)	0(0%)	0(0%)	1 (6.3%)	11 (55%)	9 (45%)	0(0%)	0(0%)	0(0%)
Lack of skills to create metadata [11]	3 (18.8%)	8 (50%)	2 (12.5%)	0 (0%)	3 (18.8%)	2 (10%)	4 (20%)	0 (0%)	7 (35%)	7 (35%)
Lack of standardised metadata [12]	4 (25%)	9 (56.3%	0 (0%)	1 (6.3%)	0 (0%)	4 (20%)	2 (10%)	5 (25%)	9 (45%)	0 (0%)
Lack of support from the university[13]	9 (56.3%)	3 (18.8%)	2 (12.5%)	2 (12.5%)	0(0%)	11 (55%)	6 (30%)	1 (5%)	2 (10%)	0(0%)
Prohibitive institutional policies [14]	0 (0%)	3 (18.8%0	13 (81.3%)	0 (0%)	0 (0%)	0 (0%)	2 (10%)	10 (50%)	3 (15%)	5 (25%)
Obsolescence of technologies[15]	1 (6.3%)	2 (12.5%)	2 (12.5%)	4 (25%)	6 (37.5%)	0 (0%)	1 (5%	0 (0%)	12 (60%)	7 (35%)
Ethical and legal norms [16]	4 (25%)	2 (12.5%)	4 (25%)	5 (31.3%)	1 (6.3%)	4 (20%)	2 (10%)	5 (25%)	3 (15%)	0 (0%)

ANOVA

[1]($F(1,34) = .301$, $p = .587$); [2]($F(1,34) = 22.027$, $p < .001$); [3]($F(1,34) = 12.200$, $p = .001$); [4]($F(1,34) = .854$, $p = .362$); [5]($F(1,34) = 2.605$, $p = .116$); [6]($F(1,34) = 5.594$, $p = .024$); [7]($F(1,34) = 3.789$, $p = .060$); [8]($F(1,34) = 4.765$, $p = .036$); [9]($F(1,34) = 1.212$, $p = .279$); [10]($F(1,34) = .036$, $p = .851$); [11]($F(1,34) = .538$, $p = .468$); [12]($F(1,34) = 2.745$, $p = .107$); [13]($F(1,34) = .650$, $p = .426$); [14]($F(1,34) = 7.686$, $p = .009$); [15]($F(1,34) = 13.949$, $p = .001$); and [16]($F(1,34) = 2.369$, $p = .133$).

Note. Constant interval percentage = 95%

Data presented in Table 17 indicates that 13 (81.2%) UNI2 respondents said failure by researchers to engage them in data curation limit their involvement in research data management whereas two (12.5%) agreed somewhat and only one (6.3%) disagreed strongly; for UNI1 staff, 13 (65%) agreed strongly, four (20%) agreed somewhat, two (10%) were neutral and one (5%) disagreed strongly. Only one (6.3%) and two (12.5%) UNI2 staff agreed strongly and agreed somewhat respectively that lack of incentives was a limitation in their involvement in research data management activities, three (18.8%) %) were neutral and 10 (62.5%) strongly disagreed; on the part of UNI1, one (5%) agreed strongly, one (5%) agreed somewhat, 11 (55%) were neutral, three (15%) disagreed somewhat and four (20%) disagreed strongly. In terms of the factor of large amounts of data to curate, one (6.3) library UNI2 staff agreed strongly, and 15 disagreed strongly and similarly, at UNI1, 11 (55%) disagreed somewhat and nine (45%) disagreed strongly. On the aspect of lack of time, one (6.3%) UNI2 staff agreed strongly, six (37.5%) disagreed somewhat and nine (56.3%) disagreed strongly; only one (5%) UNI1 staff was neutral, eight (40%) disagreed somewhat and 11 (55%) disagreed strongly. One (6.3%) UNI2 staff agreed strongly with the aspect of insufficient storage and network infrastructure, four (25%) agreed somewhat, four (25%) were neutral, and seven (43.8%) disagreed somehow; for UNI1, 12 (60%) agreed strongly, six (30%) agreed somehow and two (10%) were neutral.

On the lack of curation tools and software, seven (43.8) UNI2 staff agreed strongly, five (31.3%) agreed somehow and four (25%) were neutral; at UNI1, five (25%) agreed strongly, one (5%) agreed somewhat, 12 (60%) were neutral and two (10%) disagreed somehow. Nine (56.3%) UNI2 staff agreed strongly that there was a lack of policy frameworks and seven (43.8%) agreed somewhat; at UNI1, nine (45%) agreed strongly, five (25%) agreed somewhat, two (10%) were neutral, two (10%) disagreed somewhat and another two (20%) disagreed somewhat. On the factor of lack of curation skills and training, three (18.8%) UNI2 staff agreed strongly, eight (50%) agreed somewhat, two (12.5%) disagreed somewhat and three (18.8%) disagreed strongly; for UNI1 staff, 15 (75%) agreed strongly, one (5%) agreed somewhat and four (20%) disagreed somewhat. One (6.3%) UNI2 staff agreed strongly about lack of guidance and support, nine (56.3%) agreed somewhat, three (18.8%) were neutral and another three (18.8%) disagreed strongly; at UNI1, six (30%) agreed strongly, seven agreed somewhat, five (25%) were neutral and two (10%) disagreed somewhat. On the difficulty in finding and accessing data, 11 (68.8%) UNI2 staff agreed strongly, four (25%) agreed somewhat and one

(6.3%) disagreed strongly; similarly, at UNI1, 11 (55%) agreed strongly and nine (45%) agreed somewhat.

Three (18.8%) UNI2 staff agreed strongly about lack of skills to create metadata, eight (50%) agreed somewhat, two (12.5%) were neutral and three (18.5%) disagreed strongly; at UNI1, two (10%) agreed strongly, four (20%) agreed somewhat, seven (35%) disagreed somewhat and another seven (35%) disagreed strongly. On the aspect of lack of standardised metadata, four (25%) agreed strongly, nine (56.3%) disagreed somewhat and one (6.3%) disagreed somewhat; for UNI1 staff, four (20%) agreed strongly, two (10%) agreed somewhat, five (25%) were neutral and nine (45%) disagreed somewhat. On the factor of lack of support from the university, nine (56.3%) UNI2 staff agreed strongly, three (18.8%) agreed somewhat, two (12.5%) were neutral and another two (12.5%) disagreed somewhat; at UNI1, 11 (55%) agreed strongly, six (30%) agreed somewhat, one (5%) was neutral and two (10%) disagreed somewhat. On the factor of prohibitive institutional policies, three (18.8%) UNI2 staff agreed strongly, 13 (81.8%) %) were neutral while at UNI1, two 10 agreed somewhat, 10 (50%) were neutral, three (15%) disagreed somewhat and five (25%) disagreed strongly. On the factor of obsolescence of technologies, one (6.3%) UNI2 agreed strongly, two (12.5%) agreed somewhat, two (12.5%) were neutral, four (25%) disagreed somewhat and six (37.5%) disagreed strongly; at UNI1, only one (5%) agreed somewhat, 12 (60%) disagreed somewhat and seven (35%) disagreed strongly. The final factor was ethical and legal norms and four (25%) UNI2 staff agreed strongly, two (12.5%) agreed somewhat, four (25%) were neutral, five (31.3%) disagreed somewhat and one (6.3%) disagreed strongly; for UNI1, four (20%) agreed strongly, two (10%) agreed somewhat, five (25%) were neutral and three (15%) disagreed somewhat.

Analysis of the findings show that generally, the key factors that affected research data management at the two universities include failure by researchers to engage librarians in data curation; insufficient storage and network infrastructure; lack of policy frameworks; lack of curation skills and training; difficulty in finding and accessing data; and lack of support from the university.

ANOVA was computed to explore statistically differences in challenges in research data management between the two universities. Analysis of the *F* value and *p*-value shows that there were statistically significant differences ($p \leq 0.5$) in six factors namely; lack of incentives to curate data ($F(1,34) = 22.027$, $p < .001$); larger amounts of data to curate ($F(1,34) = 12.200$, p

= .001); lack of curation tools and software ($F(1,34) = 5.594$, $p = .024$); lack of curation skills and training ($F(1,34) = 4.765$, $p = .036$); prohibitive institutional policies ($F(1,34) = 7.686$, $p = .009$); and obsolescence of technologies ($F(1,34) = 13.949$, $p = .001$). However, a further analysis of the F value and the p-value show that there were no statistically significant differences ($p > 0.5$) in 10 factors namely; failure by researchers to engage librarians in data curation ($F(1,34) = .301$, $p = .587$); lack of time ($F(1,34) = .854$, $p = .362$); insufficient storage and network infrastructure ($F(1,34) = 2.605$, $p = .116$); lack of policy frameworks ($F(1,34) = 3.789$, $p = .060$); lack of guidance and support ($F(1,34) = 1.212$, p = .279); difficulty in accessing data ($F(1,34) = .036$, $p = .851$); lack of skills to create metadata ($F(1,34) = .538$, $p = .468$); lack of standardised metadata ($F(1,34) = 2.745$, $p = .107$); lack of support from the university ($F(1,34) = .650$, $p = .426$); and Ethical and legal norms ($F(1,34) = 2.369$, $p = .133$).

5.3. Presentation of quantitative data from researchers

This section aims at presenting and analysing data derived from researchers through the use of a questionnaire. A questionnaire was sent to 277 lecturers. Of these, 187 responded achieving a response rate of 67.5%. There is no consensus in the literature about what constitutes an acceptable response rate in social sciences research; however, Neuman (2000) and Babbie and Mouton (2001) are of the view that a response of at least 50% is acceptable for analysis. Therefore, a response rate of 67.5% achieved in the present study can be considered as adequate.

5.3.1. Demographic data

In this section, the researcher presents findings about personal information of respondents focusing on university affiliation, faculty, gender, highest qualification and the rank they held in the university.

5.3.1.1. Respondent by university

Of the 187 respondents who answered the questionnaire, 103 (55.1%) were from UNI1 and 84 (44.9%) were from UNI2.

5.3.1.2. Respondents by faculty

The questionnaire was sent to all six faculties at UNI1. The Faculty of Humanities and Social Sciences registered 46 (44.7%); Health Sciences had 20 (19.4%); and Science, Technology and Innovation had 13 (12.6%); Education had 20 (19.4%); Tourism, Hospitality and Management

registered two (1.9%); and Environmental Sciences had two (1.9%) respondents. UNI2 has only three faculties and questionnaires were sent to all of them. Results show that the Biomedical Science and Health Profession had 39 (46.4%); Medicine had 29 (34.5%); and Public Health and Family Medicine had 16 (19%). Results are presented in Figure 9.

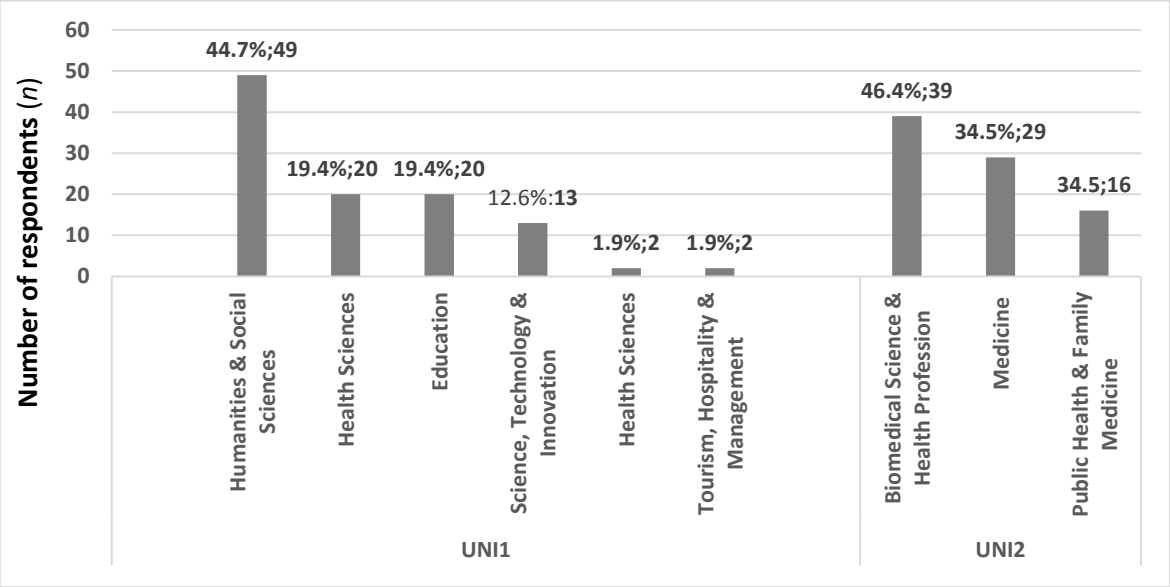


Figure 9. Respondents by faculty (N=187)

5.3.1.3. Respondents by gender, qualification and rank

Respondents were asked to indicate their gender, highest qualification and the rank they held at the time of data collection. An independent t-test was computed to determine differences between the two universities in relation to gender, qualification and rank. Findings are presented in Table 18.

Table 18. Gender, qualification and rank of respondents (N=187)

Dimension	Institution		*f*&%	Mean	SD	t-Value	Sig.
Gender	UNI1	Male	76(73.8%)	1.2621	.44195	-.536	.592**
		Female	27(26.2%)				
		Total	103(100%)				
	UNI2	Male	59(70.2%)	1.2976	.45996		
		Female	27(26.2%)				
		Total	84(100%)				
Qualification	UNI1	Masters	68 (66%)	1.3883	.58129	-3.191	.002*
		PhD	33(32%)				
		Post PhD	2(1.9%)				
		Total	103(100%)				
	UNI2	Masters	34(40.5%)	1.6667	.60785		
		PhD	44(52.4%)				
		Post PhD	6(7.1%)				
		Total	84(100%)				
Rank	UNI1	Lecturer	66(64.1%)	1.4369	.68126	-3.838	.000*
		Senior Lecturer	31(30.1%)				
		Associate Professor	5(4.9%)				
		Professor	0(0%)				
		Senior Professor	1(1%)				
		Total	103(100%)				
	UNI2	Lecturer	37(44%)	1.9167	1.02037		
		Senior Lecturer	27(32.1%)				
		Associate Professor	15(17.85)				
		Professor	5(5.95%)				
		Senior Professor	0(0%)				
		Total	84(100%)				

Note. *($p \leq 0.5$), **($p > 0.5$), constant interval percentage = 95%

Findings presented in Table 18 reveal that in terms of gender, there were 76 (73.7%) males and 27 (26.2%) females at UNI1 whereas at UNI2, there were 59 (70.2%) males and 25 (29.8%) female respondents. Results show further that there were no statistically significance differences in gender between UNI1 and UNI2 as evidenced by analysis of the t-test results that showed means (M = 1.2621*vs*. 1.2976), standard deviations (SD = .44195 *vs*. .45996), and $t(185) = -.536$, $p = .592$. This result on gender is similar to those reported under section 5.5.1.2 where it was revealed that there were more male library staff and females. Pertaining to qualification, 68 (66%) had a Masters at UNI1, 33(32%) had PhDs and two (1.9%) had Post PhDs. At UNI2, 34 (40.5%) respondents had a Masters, 44 (52.4%) had PhDs and six (7.1%) had Post PhDs. Analysis of t-tests results show that there were significant differences in qualification between UNI1 and UNI2 (M= 1.3883 *vs*. 1.6667, SD = .58129 *vs*. .60785, $t(185) = -3.191$, $p = .002$). In terms of rank, results presented in Table 18 reveal that there were 66 (64.1%) UNI1 respondents at lecturer level, 31(30.1%) at senior lecturer, five (4.9%) at

Associate Professor and one (1%) senior professor. At UNI2, there were 37 (44%) respondents at lecturer grade, 27 (32.1%) senior lecturers, 15 (11.9%) associate professors, and five (5.95%) professors. The t-test results show that there were statistically significant differences in rank between UNI1 and UNI2 (M = 1.4369 *vs*. 1.9167, SD = .68126 *vs*. 1.02037, $t(185) = -3.838$, $p < .001$).

5.3.2. Research data creation, sharing and re-use practices

Section B of the questionnaire aimed at investigating research data creation, sharing and re-use practices in public universities in Malawi. Specifically, various questionnaire items in this section helped to gather this data with a focus on research output by researchers, data format(s), factors that motivate researchers to share data, factors that affect researchers from sharing research data and data re-use practices by researchers.

5.3.2.1. Research output

This item required researchers to indicate quantities of their research output in the past 10 years in terms of research published, research papers in review, research in progress and commissioned reports. Findings are presented in Table 19.

Table 19. Cross-tabulation of research output by university (N=187)

Output	Institution	1-5	6-10	11-15	16-20	21+	None
Published	UNI1	43(41.7%)	24(23.3%)	8(7.8%)	18(17.5%)	1(1.0%)	9(8.7%)
	UNI2	53(63.1%)	12(14.3%)	6(7.1%)	11(13.1%)	2(2.4%)	0 (0%)
In review	UNI1	80(77.7%)	4(3.9%)	1(1.0%)	0 (%)	0(0%)	18(17.5%)
	UNI2	67(79.8%)	17(20.2%)	0(0%)	0(0%)	0(0%)	0(0%)
In progress	UNI1	61(59.2%)	7(6.8%)	0(0%)	4(3.9%)	0(0%)	31(30.1%)
	UNI2	67(79.8%)	11(13.1%)	0(0%)	0(0%)	3(3.6%)	3(3.6%)
Commissioned	UNI1	1(1.0%)	7(6.8%)	4(3.9%)	0(0%)	0(0%)	91(88.3%)
	UNI2	64(76.2%)	0(0%)	0(0%)	0(0%)	0(0%)	20(23.8%)

Findings presented in Table 19 show that in terms of research already published, 43 (41.7%) UNI1 researchers had one to five, 24 (23.3%) were in the range of six to 10, eight (7.8%) had 11 to 15, 18 (17%) had 21 and above and nine (8.7%) had none; on the same dimension, 53 (63.1%) UNI2 researchers had between one and five, and 12 (14.3%) had between six and 10, six (7.1%) had between 11 and 15, 11 (13.1%) were in the range of 16 to 20 and two (2.4%) had more than 21 publications. For papers in review, 80 (77.7%) UNI1 respondents were in the category of one to five, four (3.9%) were in the range of six to 10, one (1%) was in the range of 11 to 15, and 18 (17.5%) had none; for UNI2 researchers, 67 (79.8%) were in the range of

one to five and 17 (20.2%) were in the range of six to 10. For research in progress, 61 (59.2%) UNI1 researchers were in the category of one to five, seven (6.8%) were in the category of six to 10, four were in the category of 16 to 20 and 31 (30.1%) had none; at UNI2, 67 (79.8%) were in the range of one to five, 11 (13.1%) in the range six to 10, three (3.6%) had 21 and above and three (3.6%) had none. Finally, on commissioned reports, one (1%) UNI1 researcher had between one and five, seven (6.8%) had between six and 10, four had between 11 and 15 and 91 (88.3%) had none; 64 (76.2%) UNI2 researches were in the category of one to five and 20 (23.8%) did not respond.

5.3.2.2. Data formats

The respondents were asked to state the data formats in which they generated research data in the process of conducting research. An independent t-test was performed to determine differences in data formats between the two universities. Results are presented in Table 20.

Table 20. Cross-tabulation of data formats by university (N= 187)

Data format	Institution	*f*&%	Mean	SD	t-Value	Sig.
Digital text	UNI1	81(78.6%)	1.2136	.41185	3.662	.000*
	UNI2	81(96.4%)	1.0357	.18669		
Digital images	UNI1	38(36.9)	1.6311	.48487	6.245	.000
	UNI2	66(78.6%)	1.2143	.41279		
Audio recordings	UNI1	64(62.1%)	1.3786	.48742	-1.017	.311**
	UNI2	46(54.8%)	1.4524	.50072		
Video recordings	UNI1	0(%)	2.0000	.00000	a	a
	UNI2	0(%)	2.0000	.00000		
Spreadsheets	UNI1	35 (34%)	1.6602	.47596	6.306	.000*
	UNI2	64(76.2%)	1.2381	.42848		
Digital databases e.g. surveys	UNI1	56(54.4%)	1.4563	.50052	-6.969	.000*
	UNI2	9(10.7%)	1.8929	.31115		
Computer codes	UNI1	4(3.9%)	1.8544	.35446	-2.566	.011*
	UNI2	0(%)	1.9643	.18669		
Specimens	UNI1	4(3.9%)	1.9612	.19415	32.986	.000*
	UNI2	81(96.4%)	1.0357	.18669		
Spatial data	UNI1	0(%)	2.0000	.00000	3.918	.000*
	UNI2	11(13.1%)	1.8690	.33937		
Artistic products	UNI1	0(%)	2.0000	.00000	a	a
	UNI2	0(%)	2.0000	.00000		

Note. [a] t was not computed because the standard deviations of both groups were 0, *($p \leq 0.5$), ** ($p > 0.5$), constant interval percentage = 95%

Results presented in Table 20 show that 81 (78.6%) UNI1 researchers produced research data in the form of digital text, 38 (36.9%) in digital images, 64 (62.1%) in audio recordings, 35 (34%) in spreadsheets, 56 (54.4%) in digital databases, four (3.9%) in computer codes and four (3.9%) in specimens. None of UNI1 researchers generated data in the form of video recordings, spatial data and artistic products. Results from UNI2 as presented in Table 20 show that 81

(96.4%) researchers generated data in form of digital text, 66 (78.6%) in digital images, 66 (78.6%) in audio recordings, 64 (76.2%) in spreadsheets, nine (10.7%) in digital databases, 81 (96.4%) in specimens, 11 (13.1%) in spatial data. Not a single UNI2 researcher generated data in form of video recordings, computer codes and artistic products.

The t-test results also show that there were statistically significant differences (p = < 0.5) between UNI1 and UNI2 across all dimensions as it can be observed about the minimum mean (M = 1.0357) and the maximum mean (M = 2.0000) and likewise, the minimum standard deviation (SD = .00000) and the maximum standard deviation (SD .50072). In more specific terms, an analysis of the p-value revealed that there were statistically significant differences (p = < 0.5) in nine dimensions including digital text ($t(185) = 3.662$, $p < .001$); digital images ($t(185) = 6.245$, $p < .001$), spreadsheets ($t(185) = 6.306$, $p < .001$); digital databases ($t(185) = -6.969$, $p < .001$); computer codes ($t(185) = -2.566$, $p = .011$); specimens ($t(185) = 32.986$, $p < .001$); and spatial ($t(185) = 3.918$, $p < .001$). Analysis of the p-value revealed further that there were no statistically significance differences ($p > 0.5$) in the dimension of audio recordings ($t(185) = -1.017$, $p = .311$).

5.3.2.3. Research data sharing

Researchers were asked to state if they shared the data they generated with other researchers and other research stakeholders. Findings are presented in Figure 10.

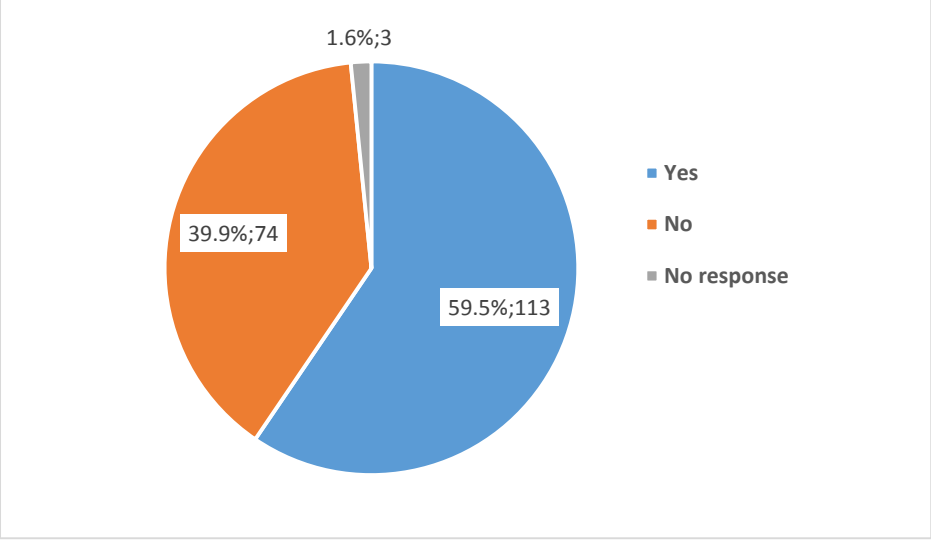


Figure 10. Research data sharing by researchers (N=187)

Findings presented in Figure 10 show that the majority of researchers with a score of 113 (59.5%) said yes, 74 (39.9%) said no while three (1.6%) did not respond. A further analysis of

results through crosstabulation showed that of the 113 respondents who said shared their data, 64 (62.1%) were UNI1 researchers while 49 (58%) were UNI2 researchers.

5.3.2.4. Factors that motivate researchers to share research data

Researchers were asked to state factors that encouraged them to share the research data they generated. Results are captured in Table 21.

Table 21. Factors that motivate researchers to share research data (N=187)

Factors	UNI1				UNI2			
	Yes		No		Yes		No	
	f	%	*f*	%	*f*	%	*f*	%
Journal policies	26	25.2	77	74.8	23	27.4	61	72.6
Research funders	6	5.8	97	94.2	32	38.1	52	61.9
University policy	0	0	103	100	23	27.4	61	72.6
Open Access	7	6.8%	96	93.2	6	7.1	78	92.9%
Personal initiative	63	61.2	40	38.8	49	58.3	35	41.7%

It is clear from the findings presented in Table 21 that 26 (25.2%) UNI1 researchers shared data because they were influenced by journal policies while 23 (27.4%) UNI2 researchers were influenced by the same aspect. Six (5.8 %) UNI1 researchers were influenced by research funders whereas 32 (38.1%) UNI2 researchers were influenced by the same factor. While none of the UNI1 researchers was influenced by university policy, 23 (27.4%) UNI2 researchers were compelled by this aspect. The open access factor influenced seven (6.8%) UNI1 researchers to share data and similarly, six (7.1%) UNI2 researchers were compelled by this factor. Finally, the factor of personal initiative influenced 63 (61.2%) UNI1 researchers and 49 (58.3%) UNI2 researchers to share the data they generated. These results suggest that researchers at both universities were mainly compelled to share data by personal initiatives followed by journal policies.

5.3.2.5. Data sharing tools

In this item, respondents were provided with a list of various data sharing tools and were asked to indicate the extent to which they used them in sharing the research data they generated. The item made use of a four-point Likert Scale ranging from *all*, *most*, *some,* and *none*. Researchers who selected *all* meant that they shared all their data using that particular data sharing tool, those who selected *most* meant that they shared most of their data using that particular data sharing tool, those that selected *some* meant that they shared some of their data using that particular data sharing tool, and those that selected *none* meant they never used that particular

data sharing tool to share their data. Findings are shown in Table 22. ANOVA was computed to explore differences in research data sharing tools by university affiliation.

Table 22. Data sharing tools by researchers (N=187)

Data sharing tools	Institution									
	UNI1 (*n*=103)					UNI2 (*n*=84)				
	All	Most	Some	None	No response	All	Most	Some	None	No response
External drives[1]	14 (13.6%)	20 (19.4%)	14 13.6%	25 (24.3%)	30 (29.1%)	14 (16.7%)	20 (23.8%)	9 (10.7%)	6 (7.1%)	35 41.7%
Emails[2]	20 (19.4%)	9 (8.7%)	32 (31.1%)	16 (15.5%)	16 15.5%	20 (23.8%)	23 (27.4%)	3 (3.6%)	3 3.6%	35 (41.7%)
e-journals' websites[3]	0(%)	0(0%)	7 (6.8%)	68 (66%)	68 (66%)	17 (20.2%)	6 (7.1%)	0 (0%)	26 (31%)	35 (41.7%)
Social networks[4]	0(%)	0(0%)	0(%)	75(72.8 %)	28 (27.2%)	0(0%)	0(0%)	0(0%)	49 58.3%	35 41.7%
Blogs/wikis[5]	0(%)	0(%)	0(%)	75 (72.8%)	28 (27.2%)	0(0%)	0(0%)	3 (3.6%)	46 (54.8%)	35 (41.7%)
Clouds[6]	0(%)	0(%)	20 (19.4%)	55 (53.4%)	28 (27.2%)	0(%)	14 (16.7%)	18 (21.4%)	17 (20.2%)	35 (41.7%)
University repositories[7]	18 17.5%	4 (3.9%)	14 (13.6%)	40 (38.8%)	27 (26.2%)	3 (3.6%)	0(0%)	12 (14.3%)	34 (40.5%)	35 (41.7%)
Funders websites[8]	0(0%)	0(0%)	0(0%)	71 (68.9%)	32 31.1%	3 (3.6%)	0(0%)	9 (10.7%)	37 (44%)	35 (41.7%)
University website[9]	0(0%)	0(0%)	0(0%)	71 (68.9%	32 (31.1%)	0(0%)	0(0%)	0(0%)	49(58.3 %)	35 (41.7%)
Principal investigator's website[10]	0(0%)	0(0%)	0(0%)	0(0%)	103 (100%)	0(0%)	0(0%)	0(0%)	0(0%)	84 (100%)
National network[11]	0(0%)	0(0%)	0(0%)	0(0%)	103 (100%)	0(0%)	0(0%)	0(0%)	0(0%)	84 (100%)
Regional network[12]	0(0%)	0(0%)	0(0%)	71 (68.9%)	32 (31.1%)	0(0%)	0(0%)	3 (3.6%)	46 (54.8%)	35 (41.7%)
Global network[13]	0(0%)	0(0%)	13 (12.6%)	58 (56.3%)	32 (31.1%)	0(0%)	0(0%)	9 (10.7%)	40 (47.6%)	35 (41.7%)

ANOVA

[1]($F(1,185) = .014$, $p = .907$); [2]($F(1,185) = .081$, $p = .776$); [3]($F(1,185) = 2.334$, $p = .128$); [4]($F(1,185) = 14.280$, $p < .001$); [5]($F(1,185) = 2.206$, $p = .139$); [6]($F(1,185) = 1.009$, $p = .316$); [7]($F(1,185) = 13.212$, $p < .001$); [8]($F(1,185) = 1.176$, $p = .280$);[9] ($F(1,185) = 1.176$, $p = .280$); [10] ($F(1,185) = 4.518$, $p = .035$); [11]($F(1,185) = 4.518$, $p = .035$); [12]($F(1,185) = 8.143$, $p = .005$); and [13] ($F(1,185) = 4.518$, $p = .035$).

Note. Constant interval percentage = 95%

Table 22 shows the tools that researchers used to share the research data they generated. For UNI1 researchers, it is clear that some used external hard drives with 14 (13.6%) indicating all, 20 (19.4%) indicating most and 14 (13.6%) indicating some, 25 (24.3%) indicating none and 30 (29.1%) did not respond. Emails were another form of a sharing tool used by UNI1 researchers with 20 (19.4%) indicating all, nine (8.7%) indicating most, 32 (31.1%) indicating some, 16 (15.5%) said none and another 16 (15.5%) did not respond. Eighteen 18 (17.5%)

UNI1 researchers indicated they shared all data using university repositories, four (3.9%) said most, 14 (13.6%) indicated some, 40 (38.8%) indicated none and 27 (26.2%) did not respond. A small number of researchers shared some of their data using clouds, journal websites and global network with scores of 20 (19.4%), seven (6.8%) and 13 (12.6%) respectively. On the other hand, results presented in Table 22 show that none of the UNI1 researchers used social networks, blogs/wikis, funder's websites, university websites, a principal investigator's website, national network, or regional network.

Results presented in Table 22 show that for UNI2 researchers, 14 (16.7%) indicated they shared all their research data using external hard drives, 20 (23.8%) said most, nine (10.7%) said some, six (7.1%) said none and 35 (41.7%) did not respond. Twenty (23.8%) used emails to share all their data, 23 (27.4%) indicated most, three (3.6%) said some, another three (3.6%) said none and 35 (41.7%) did not respond. Journal websites were used by 17 (20.2%) researchers to share all their data, six (7.1%) indicated most, 26 (31%) said none and 35 (41.7%) did not respond. A smaller number of UNI2 researchers used clouds with a score of 32 (38.1%) who indicated most/some; university repositories with a score of 15 (17.9%) indicating all/some; 12 (14.3%) indicating all/some use of funders' websites; regional networks with a score of three (3.6%) indicating some; and global networks with a score of nine (10.7%) indicating some. Results show further that none of the UNI2 researchers used social networks, blogs/wikis, the university website, principal investigator's website or national network.

Analysis of the findings at both universities shows that, the common research data sharing tools include external hard drives and emails. These results support those from librarians as presented in section 5.5.3.1 in the sense that most librarians at both universities indicated that external hard drives and email accounts were provided by their universities. On the other hand, UNI1 researchers also used university repositories which UNI2 staff did not use while UNI2 researchers also used journals' websites which UNI1 researchers did not use. Researchers at both universities did not use or used social networks, blogs/wikis, journal websites, funder's websites, university websites, principal investigators' websites, a national network, and a regional network minimally.

The researcher computed ANOVA to explore differences in data sharing tools amongst respondents by university affiliation. ANOVA results revealed that there were statistically significant differences in social networks (F (1,185) = 14.280, $p < .001$), university repositories (F(1,185) = 13.212, $p < .001$), principal investigator's website (F(1,185) = 4.518, $p = .035$),

national network ($F(1,185) = 4.518$, $p = .035$), regional network ($F(1,185) = 8.143$, $p = .005$) and global network ($F(1,185) = 4.518$, $p = .035$). On the other hand, there were no statistically significance differences in some facilities including external hard drives ($F(1,185) = .014$, $p = .907$), emails ($F(1,185) = .081$, $p = .776$), e-journals' websites ($F(1,185 = 2.334$, $p = .128$), Blogs/wikis ($F(1,185) = 2.206$, $p = .139$), clouds ($F(1,185) = 1.009$, $p = .316$), funders websites ($F(1,185) = 1.176$, $p = .280$), and university websites ($F(1,185) = 1.176$, $p = .280$).

5.3.2.6. Factors that discourage researchers from sharing research data

The aim of this item was to determine the extent to which various factors discouraged researchers from sharing research data they generated with other researchers. The question was in form of a Lickert scale and researchers were required to indicate the extent to which each factor affected them by choosing from the options of *Agree Strongly, Agree Somewhat, Neutral, Disagree Somewhat* and *Disagree Strongly.* Findings are shown in Table 23. ANOVA was also computed to explore differences in factors that discourage researchers from sharing data by university affiliation.

Table 23. Factors that discourage researchers from sharing their research data (N=187)

Factors	Institution									
	UNI1 (*n*=103)					UNI2 (*n*=84)				
	Agree Strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Lack of incentives[1]	59 (57.3%)	37 (35.9%)	0(0%)	4 3.9%	3 (2.9%)	0(0%)	3 (3.6%)	0(0%)	29 (34.5%)	49 (58.3%)
Lack of funding[2]	26 (25.2%)	43 (41.7%)	19 (18.4%)	15 14.6%	0(0%)	0(0%)	6 (7.1%)	0(0%)	58 (69.0%)	17 (20.2%)
Lack of standards or guidelines[3]	40 (38.8%)	15 (14.6%)	19 (18.4%)	7 (6.8%)	22 (21.4%)	15 (14.6%)	14 (16.7%)	12 (14.3%)	6 (7.1%)	11 (13.1%)
Data is not fully documented[4]	44 (42.7%)	37 (35.9%)	10 (9.7%)	7 (6.8%)	5 (4.9%)	58 (69.0%)	17 (20.2%)	6 (7.1%)	3 (3.6%)	0(0%)
No place to put the data[5]	39 (37.9%)	38 (36.9%)	13 (12.6%)	2 (1.9%)	11 (10.7%)	0(0%)	0(0%)	3 (3.6%)	61 (72.6%)	20 (23.8%)
License agreements[6]	0(0%)	2 (1.9%)	13 (12.6%)	49 (47.6%)	39 (37.9%)	0(0%)	3 (3.6%)	0(0%)	28 (33.3%)	53 (63.1%)
Lose control over my data[7]	1 (1.0%)	0(0%)	0(0%)	51 (49.5%)	50 (48.5%)	61 (72.6%)	17 (20.2%)	6 (7.1%)	0(0%)	0(0%)
Insufficient skills[8]	59 (57.3%)	25 (24.3%)	15 (14.6%)	0(0%)	4 (3.9%)	58 (69.0%)	26 (31.0%)	0(0%)	0(0%)	0(0%)
Data format is not widely readable[9]	37 (35.9%)	22 21.4%	19 (18.4%)	22 21.4%	3 (2.9%)	26 (31.0%)	55 (65.5%)	0(0%)	3 (3.6%)	0(0%)
Data may be misinterpreted[10]	67 (65.0%)	8 (7.8%)	10 (9.7%)	13 (12.6%)	4 (3.9%)	0(0%)	0(0%)	3 (3.6%)	32 (38.1%)	49 (58.3%)
University owns data[11]	0(0%)	0(0%)	0(0%)	46 (44.7%)	56 (54.4%)	3 (3.6%)	0(0%)	0(0%)	69 (82.1%)	12 (14.3%)
Funding agency owns data[12]	3 (2.9%)	9 (8.7%)	20 (19.4%)	42 (40.8%)	28 (27.2%)	43 (51.2%)	41 (48.8%)	0(0%)	0(0%)	0(0%)
Insufficient time[13]	24 (23.3%)	52 (50.5%)	0(0%)	18 17.5%	8 7.8%	78 (92.9%)	6 (7.1%)	0(0%)	0(0%)	0(0%)

ANOVA

[1]($F(1,185) = 592.301$, $p < .001$); [2]($F(1,185) = 12.039$, $p = .001$); [3]($F(1,185) = 20.066$, $p < .001$); [4]($F(1,185) = 39.126$, $p < .001$); [5]($F(1,185) = 211.905$, $p < .001$); [6]($F(1,185) = 308.312$, $p < .001$); [7]($F(1,185) = 1186.836$, $p < .001$); [8]($F(1,185) = 10.462$, $p = .001$); [9]($F(1,185) = 106.296$, $p < .001$); [10]($F(1,185) = 301.942$, $p < .001$); [11]($F(1,185) = 35.945$, $p < .001$); [12]($F(1,185) = 157.276$, $p < .001$); [13]($F(1,185) = 157.580$, $p < .001$).

Note. Constant interval percentage = 95%

Table 23 shows factors that discouraged researchers from sharing the data they generated. The key factors that affected UNI1 researchers include lack of incentives where 59 (57.3%) agreed strongly, 37 (35.9%) agreed somewhat; four (3.9%) disagreed somewhat and three (2.9%) disagreed strongly; lack of funding where 26 (25.2%) agreed strongly, 43 (41.7%) agreed somewhat, 19 (18.4%) were neutral and 15 (14.6%) disagreed somewhat; lack of standards where 40 (38.8%) agreed strongly, 15 (14.6%) agreed somewhat, 19 (18.4%) were neutral, seven (6.8%) disagreed somewhat and 22 (21.4%) disagreed strongly; data not fully

documented with 44 (42.7%) who agreed strongly, 37 (35.9%) agreed somewhat; 10 (9.7%) were neutral, seven (6.8%) disagreed somewhat and five (4.9%) disagreed strongly; and lack of storage facilities with 39 (37.9%) who agreed strongly, 38 (36.9%) who agreed somewhat, 13 (12.6%) were neutral, two (1.9%) disagreed somewhat and 11 (10.7%) disagreed strongly. Fifty nine (57.3%) UNI1 staff agreed strongly, 25 (24.3%) agreed somewhat, 15 (14.6%) were neutral and four (3.9%) disagreed strongly with the factor of insufficient skills; 37 (35.9%) agreed strongly, 22 (21.4%) agreed somewhat, 19 (18.4%) were neutral, 22 (21.4%) disagreed somewhat and three (2.9%) disagreed strongly with the factor of data format not widely readable; 67 (65%) agreed strongly, eight (7.8%) agreed somewhat, 10 (9.7%) were neutral, 13 (12.6%) disagreed somewhat and four (3.9%) disagreed strongly with the factor their data may be misinterpreted; and 24 (23.3%) agreed strongly, 52 (50.5%) agreed somewhat, 18 (17.5%) disagreed somewhat and eight 7.8% that they had insufficient time.

For UNI2 staff, findings presented in Table 23 show that 15 (14.6%) agreed strongly about lack of standards, 14 (16.7%) agreed somewhat, 12 (14.3%) were neutral, six (7.1%) disagreed somewhat and 11 (13.1%) disagreed strongly; 58 (69%) strongly agreed, 17 (20.2%) agreed somewhat, six (7.1%) were neutral, three (3.6%) disagreed somewhat that data is not fully documented; 61(72.6%) agreed strongly, 17 (20.2%) agreed somewhat and six (7.1%) were neutral about the factor of losing control over their data; 58 (69%) agreed strongly and 26 (31%) agreed somewhat that they had insufficient skills; 26 (31%) agreed strongly, 55(65.5%) agreed somewhat and three (3.6%) disagreed somewhat that their data format was not widely readable; and 43 (51.2%) UNI2 staff agreed strongly and 41 (48.8%) agreed somewhat that funding agencies owned the data they generated; and 78 (92.9%) agreed strongly and six (7.1%) agreed somewhat that they lacked time for sharing the data.

Analysis of the results shows that the factors that affected researchers at both universities include data not fully documented, insufficient skills, lack of standards or guidelines, data format not widely readable and insufficient time to share data. Factors applicable to UNI1 only included lack of incentives, lack of funding, data may be misinterpreted, and unavailability of storage facilities. Factors limited to UNI2 include loss of control over data and data being owned by funding agencies. Two factors of namely, university owns data and data licence issues did not discourage researchers from sharing their research at both universities.

The researcher computed ANOVA to determine if there were statistically significant differences in factors that discouraged respondents by university affiliation. Results showed

statistically significant differences in all dimensions ($p < 0.5$) as follows: Lack of incentives ($F(1,185) = 592.301$, $p < .001$); lack of funding ($F(1,185) = 12.039$, $p = .001$); lack of standards or guidelines ($F(1,185) = 20.066$, $p < .001$); data is not fully documented ($F(1,185) = 39.126$, $p < .001$), no place to put the data ($F(1,185) = 211.905$, $p < .001$); license agreements ($F(1,185) = 308.312$, $p < .001$); lose control over my data ($F(1,185) = 1186.836$, $p < .001$); insufficient skills ($F(1,185) = 10.462$, $p = .001$); data format is not widely readable ($F(1,185) = 106.296$, $p < .001$), data may be misinterpreted ($F(1,185) = 301.942$, $p < .001$); data may be misinterpreted ($F(1,185) = 301.942$, $p < .001$); university owns data ($F(1,185) = 35.945$, $p < .001$); funding agency owns data ($F(1,185) = 157.276$, $p < .001$); and insufficient time ($F(1,185) = 157.580$, $p < .001$).

5.3.2.7. Conditions for sharing research data

The researcher presented researchers with various factors and asked them to indicate the extent to which these factors would encourage them to share the research data they generated. Findings are presented in Table 24. ANOVA was computed to explore differences in conditions that encourage research data between the two universities.

Table 24. Conditions that encourage researchers to share data (N=187)

Factors	Institution									
	UNI1 (*n*=103)					UNI2 (*n*=84)				
	Agree Strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Place some data in central repositories without restrictions[1]	0 (0.0%)	15 (14.6%)	15 (14.6%)	59 (57.3%)	29 (28.2%)	0 (0.0%)	3 (3.6%)	9 (10.7%)	17 (20.2%)	55 (65.5%)
Place all data in central repository without restrictions[2]	2 (1.9%)	7 (6.8%)	0 (0.0%)	27 (26.2%)	55 (53.4%)	0 (0.0%)	0 (0.0%)	9 (10.7%)	55 (65.5%)	20 (23.8%)
More likely to make my data available if I place conditions on access[3]	64 (62.1%)	26 (25.2%)	7 (6.8%)	6 (5.8%)	0 (0.0%)	81 (96.4%)	3 (3.6%)	0(0.0%)	0(0.0%)	0 (0.0%)
Data should be cited when used by other researchers[4]	32 (31.1%)	63 (61.2%)	8 (7.8%)	0 (0.0%)	0 (0.0%)	23 (27.4%)	58 (69.0%)	3 (3.6%)	0 (0.0%)	0 (0.0%)
Create new data sets from shared data[5]	43 (41.7%)	43 (41.7%)	0 (0.0%)	22 (21.4%)	0 (0.0%)	37 (44.0%)	41 (48.9%)	6 (7.1%)	0 (0.0%)	0 (0.0%)

ANOVA

[1](F(1,185) = 30.146, p < .001); [2](F(1,185) = .001, p = .975); [3](F(1,185) = 30.432, p < .001); [4](F(1,185) = 51.647, p < .001); [5](F(1,185) = 7.554, p = .007).

Note. Constant interval percentage = 95%

Findings presented in Table 24 reveal that 64 (621%) UNI1 researchers agreed strongly, 26 (25.2%) agreed somewhat, seven (6.8%) were neutral and six (5.8%) disagreed somewhat on the condition of placing restrictions on access to the data they shared; and similarly, 81 (96.4%) UNI2 researchers agreed strongly and three (3.6%) agreed somewhat on the same condition. On the condition that data should be cited when used by other researchers, 32 (31.1%) UNI1 researchers agreed strongly, 63 (61.2%) agreed somewhat and eight (7.8%) were neutral; and likewise, 23 (27.4%) UNI2 researchers agreed strongly, 58 (69%) agreed somewhat and three (3.6%) were neutral about the same condition. Finally, on the factors of creating new data sets from shared data, 43 (41.7%) UNI1 researchers agreed strongly, 43 (41.7%) agreed somewhat and 22 (21.4%) disagreed somewhat; and 37 (44%) UNI2 staff agreed strongly, 41 (48.9%) agreed somewhat while six (7.1%) were neutral on the need by users to create new data sets from shared data.

Findings presented in Table 24 reveal that there were three key conditions that could motivate researchers at both universities to share the data they generate which include placing restrictions on the data they share, the need to have their data cited by users and the need to create new data sets from the data they shared. However, the other two factors (placing some data in central repositories without restrictions and placing all data in a central repository without restrictions) did not motivate researchers to share their data.

Results were further analysed by computing the ANOVA to determine if there were significant differences in sharing conditions amongst researchers by university affiliations. Results showed that there were significant differences in four dimensions namely; place some data in central repositories without restrictions ($F(1,185) = 30.146$, $p < .001$), place restrictions on access on the data they share ($F(1,185) = 30.432$, $p < .001$); data should be cited when used by other researchers ($F(1,185) = 51.647$, $p < .001$); and need by users to create new data sets from shared data ($F(1,185) = 7.554$, $p = .007$). Only the condition of placing all data in a central repository without restrictions showed no statistically significant differences between the two universities ($F(1,185) = .001$, $p = .975$).

5.3.2.8. Research data re-use by researchers

Researchers were asked to indicate the frequency with which they used research data generated by other researchers or research institutions. Results show that 45 (23.7%) said always, 43 (22.6%) said frequently, 20 (10.5%) said occasionally, 79 (41.6%) said never and two (1.6%) did not respond to this question. The results suggest that generally, researchers used research data produced by other researchers but on small scale.

5.3.2.9. Factors that affect data re-use by researchers

Researchers were provided with a list of nine factors and were asked the extent to which each of the factors discouraged them from using data generated by other researchers or research institutions. Findings are shown in Table 25. An independent t-test was also computed to explore differences in factors that affect data re-use between the two universities.

Table 25. Factors that affect data re-use by researchers (N=187)

Factors	Institution									
	UNI1 (*n*=103)					UNI2 (*n*=84)				
	Agree Strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Difficult to find, discover, or access reusable data[1]	83 (80.6%)	20 (19.4%)	0(0%)	0(0%)	0(0%)	23 (27.4%)	61 72.6%	0(0%)	0(0%)	0(0%)
Hard to integrate with my own data[2]	2 (1.9%)	20 (19.4%)	13 (12.6%)	50 (48.5%)	18 (17.5%)	17 (20.2%)	6 (7.1%)	26 (31.0%)	35 (41.7%)	0(0%)
Not trusting others' collection methods[3]	1 (1.0%)	18 (17.5%)	7 (6.8%)	74 (71.8%)	3 (2.9%)	0 (0.0%)	0 (0.0%)	6 (7.1%)	64 (76.2%)	14 (16.7%)
Data may be misinterpreted due to its complexity	1 (1.0%)	7 (6.8%)	28 (27.2%)	38 (36.9%)	29 (28.2%)	3 (3.6%)	14 (16.7%)	0 (0.0%)	67 (79.8%)	0 (0.0%)
Lack of common or standard formats[5]	52 (50.5%)	48 (46.6%)	1 (1%)	1 (1%)	1 (1%)	75 (89.3%)	3 (3.6%)	3 (3.6%)	0 (0.0%)	3 (3.6%)
Lack of metadata[6]	42 (40.8%)	41 (39.8%)	15 (14.6%)	1 (1%)	4 (3.9%)	43 (51.2%)	41 (48.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Data may be misinterpreted due to poor quality[7]	30 (29.1%)	26 (25.2%)	15 (14.6%)	6 (5.8%)	26 (25.2%)	15 (17.9%)	55 (65.5%)	0 (0.0%)	14 (16.7%)	0 (0.0%)
Data may be used in other ways than intended[8]	2 (1.9%)	15 (14.6%)	18 (17.5%)	22 (21.4%)	46 (44.7%)	0 (0.0%)	0 (0.0%)	38 (45.2%)	34 (40.5%)	12 (14.3%)
Legal/ethical restrictions[9]	16 (15.5%)	33 (32%)	49 (47.6%)	4 (3.9%)	1 (1%)	29 (34.5%)	17 (20.2%)	0 (0.0%)	38 (45.2%)	0 (0.0%)

Independent t-test

[1](t (185) = -8.592, p < .001); [2](t (185) = 4.113, p < .001); [3](t (185) = -4.938, p < .001); [4](t (185) = -3.867, p < .001); [5](t (185) = 2.763, p = .006); [6](t (185) =3.308, p = .001); [7](t (185) = 2.978, p = .003); [8](t (185) = -1.952, p = .052); [9](t (185) = -.815, p = .416).

Note. Constant interval percentage = 95%

Findings presented in Table 25 show that 83 (80.6%) UNI1 researchers agreed strongly and 20 (19.4%) agreed somewhat with the factor of difficulty in finding, discovering or accessing reusable data and on the same aspect, 23 (27.4%) UNI2 staff agreed strongly and 61 (72.6%) agreed somewhat. On the factor of lack of common or standard formats, 52 (50.5%) UNI1 researchers agreed strongly, 48 (46.6%) agreed somewhat, one (1%) was neutral, one (1%) disagreed somewhat and another one (1%) disagreed strongly; and similarly, 75 (89.3%) UNI2 researchers agreed strongly, three (3.6%) agreed somewhat, three (3.6%) were neutral and another three (3.6%) disagreed strongly. Another key factor was lack of metadata where 42 (40.8%) UNI1 researchers agreed strongly, 41 (39.8%) agreed somewhat, 15 (14.6%) were

neutral, one (1%) disagreed somewhat and four (3.9%) disagreed strongly; and likewise, 43 (51.2%) UNI2 researchers agreed strongly and 41 (48.8%) agreed somewhat. On the factor that data may be misinterpreted due to poor quality, 30 (29.1%) UNI1 researchers agreed strongly, 26 (25.2%) agreed somewhat, 15 (14.6%) were neutral, six disagreed somewhat and 26 (25.2%) disagreed strongly; while 15 (17.9%) UNI2 researchers agreed strongly, 55 (65.5%) agreed somewhat and 14 disagreed somewhat. On the aspect of legal/ethical restrictions, 16 (15.5%) UNI1 researchers agreed strongly, 33 (32%) agreed somewhat, 49 (47.6%) were neutral, four (3.9%) disagreed somewhat and one (1%) disagreed strongly; while 29 (34.5%) UNI2 researchers agreed strongly, 17 (20.2%) agreed somewhat, and 38 (45.2%) disagreed somewhat.

From these findings, it is clear that some factors namely; hard to integrate data, not trusting other researchers' collection methods, data may be misinterpreted due to its complexity, and data may be used in other ways than intended did not discourage researchers from re-using data generated by other researchers.

The researcher wanted to determine if there were statistically significance differences in the factors that discouraged researchers to re-use data between the two universes and an independent t-test was computed. Results revealed that there were statistically significance differences in eight factors ($p < 0.5$) that include difficult to find, discover, or access reusable data ($t(185) = -8.592$, $p < .001$); difficulty to integrate with own data ($t(185) = 4.113$, $p < .001$); not trusting others' collection methods ($t(185) = -4.938$, $p < .001$), data may be misinterpreted due complexity ($t(185) = -3.867$, $p < .001$); lack of common or standard formats ($t(185) = 2.763$, $p = .006$); lack of metadata ($t(185) = 3.308$, $p = .001$); data may be misinterpreted due to poor quality ($t(185) = 2.978$, $p = .003$); data may be used in other ways than intended ($t(185) = -1.952$, $p = .052$). There were however no statistically significant differences by university affiliation ($p > 0.5$) on the factor of legal/ethical restrictions ($t(185) = -.815$, $p = .416$).

5.3.3. Research data preservation practices

Section *C* of the questionnaire (See Appendix B) aimed at investigating research data preservation practices. Specific issues that were investigated included the need for research data preservation, the lifespan of research data, quantities of research data, storage facilities, data back-up strategies, research data infrastructure, research data management training, metadata, skills for research data management, and support researchers seek from various professionals

5.3.3.1. Need for research data preservation

Researchers were asked to indicate if it was necessary to preserve research data. Findings revealed that all 187 (100%) researchers said it was necessary to preserve research data.

5.3.3.2. Lifespan of preserved research data

The researcher asked respondents to indicate the period that the data they preserved could remain valuable and accessible. Crosstabulation of the results is presented in Table 26.

Table 26. Crosstabulation of data duration by university (N=187)

Institution	Duration					
	Indefinitely	**10 – 20 years**	**5–10 years**	**3–5 years**	**1-2**	**Not sure**
UNI1	18 (17.5%)	8 (7.8%)	41(39.8%)	10 (9.7%)	0(0%)	26 (25.2%)
UNI2	9 (10.7%)	0(0%)	58 (69.0%)	0(0%)	0(0%)	17 (20.2%)

Data presented in Table 26 shows that 18 (17.5%) UNI1 researchers indicated that their data could remain valuable indefinitely, eight (7.8%) indicated five to 10 years, 10 (9.7%) said three to five years and 26 (25.2%) did not answer the question. Similarly, nine (10.7%) UNI2 researchers said their data could remain valuable indefinitely, 58 (69%) said five to 10 years and 17 (20.2%) did not respond to the question. Analysis of these results shows that the majority of researchers at both universities were of the view that their data could remain valuable for a period of between five and 10 years.

5.3.3.3. Quantity research data produced by researchers

In this item, researchers were asked to indicate the largest amount of research data that they had generated. Findings are presented in Table 27.

Table 27. Crosstabulation of the amount of data by university (N=187)

Institution	**Amount of digital research data** ***Key**: GB = Gigabyte, TB= Terabyte, PB = Petabyte*					
	1GB -100GB	**100GB -1TB**	**1TB -100TB**	**100TB -1PB**	**>1PB**	**Not sure**
UNI1	65 (63.1%)	12 (11.7%)	4 (3.9%)	7 (6.8%)	0 (0%)	15 (14.5%)
UNI2	3 (3.6%)	72 (85.7%)	6 (7.1%)	0 (0%)	0 (0%)	3 (3.6%)

An can be seen in Table 27, 65 (63.1%) UNI1 researchers had produced between one and 100 GB, 12 (11.7%) had produced between 100 GB and one TB, four (3.9%) indicated between one TB and 100 TB, seven (6.8%) said between 100 TB and one PB and 15 (14.5%) said they

were not sure. For UNI2 staff, three (3.6%) indicated between one GB and 100 GB, 72 (85.7%) said between 100 GB and one TB, six (7.1%) indicated between one TB and 100 TB, and three (3.6%) said they were not sure. Generally, findings suggest that the majority of UNI1 researchers produced between one and 100 GB of data whereas UNI2 researchers produced between 100 GB and one terabyte of data.

5.3.3.4. Digital data storage facilities

Researchers were asked to state the digital storage facilities that they used to preserve their research data. To answer this question, crosstabulation, means, standard deviations and independence t-tests were tabulated and findings are presented in Table 28.

Table 28. Digital storage devices by university (N=187)

Storage facility	University	N	*n* & (%)	M	SD	t-value	Sig.
Personal computers	UNI1	103	103 (100)	1.0000	.00000	a.	a.
	UNI2	84	84 (100)	1.0000	.00000		
Office computers	UNI1	103	63 (61.2)	1.3883	.48976	2.206	.029*
	UNI2	84	64 (76.2	1.2381	.42848		
External hard drives	UNI1	103	99 (96.1)	1.0388	.19415	1.832	.069**
	UNI2	84	84 (100)	1.0000	.00000		
CDs	UNI1	103	21 (20.4)	2.4078	2.36576	2.557	.011*
	UNI2	84	22 (26.2)	1.7381	.44231		
Institution's available networked capacity	UNI1	103	0(0)	2.0000	.00000	a.	a.
	UNI2	84	0(0)	2.0000	.00000		
Commercial software or services	UNI1	103	1 (1)	3.1553	4.71077	2.247	.026*
	UNI2	84	0 (0)	2.0000	.00000		
Freely available software or services (Google Drive)	UNI1	103	83(80.6)	1.1942	.39750	4.475	.000*
	UNI2	84	84 (100)	1.0000	.00000		
USB	UNI1	103	99 (96.1)	1.0388	.19415	1.832	.069 **
	UNI2	84	84 (100)	1.0000	.00000		
Email account(s)	UNI1	103	84 (81.6)	1.1845	.38976	4.336	.000*
	UNI2	84	84 (100)	1.0000	.00000		

Note. [a] t was not computed because the standard deviations of both groups were 0, *($p \leq 0.5$), ** ($p > 0.5$), constant interval percentage = 95%

Table 28 presents data on digital storage facilities used by researchers to store the data they generated. Findings show that all 103 (100%) UNI1 researchers use personal computers, 63 (61.2%) used office computers, 99 (96.1%) used external hard drives, 83(80.6%) used freely available software or services such as Google Drive, 99 (96.1%) used flash disks and 84 (81.6%) used email accounts. Similarly, all 84 (100%) UNI2 researchers used personal computers, 64 (76.2%) used office computers, all 84 (100%) used external hard drives, and freely available software or services such as Google Drive flash disks and all 84 (100%) used

email accounts. Results reveal that the common digital storage facilities used by researchers at both universities include personal computers, office computers external hard drives, freely available software or services such as Google Drive, flash disks and email accounts. Results showed further that CDs, institution's available networks and commercial software or services were not used by researchers. Results reported in section 5.5.3.1 also showed that flash discs, computers, email accounts, networks, free software and external hard drives were the common technological storage facilities available in libraries that could be used by researchers to manage their data.

The researcher attempted to find out if there were statistically significant differences in storage facilities by university affiliation. To achieve this objective, means, standard deviations and independence t-tests were tabulated. In two dimensions namely; personal computers and institution's available networked capacity, standard deviations were zero hence it was impossible to compute the t-value. Generally, there were statistically significant differences in data storage facilities by university affiliation. Table 28 shows that across the nine dimensions, the means range between (M = 1.0000) and (M = 3.1553) while the standard deviations range between (SD = .00000) and (SD = 4.71077). Specifically, analysis of the t-value and p-value reveals that there were statistically significant differences in five dimensions ($p > 0.5$) namely, office computers ($t(185) = 2.206$, $p = .029$); CD-ROMs ($t(185) = 2.557$, $p = .011$); commercial software or services ($t(185) = 2.247$, $p = .026$); freely available software or services ($t(185) = 4.475$, $p < .001$); and email accounts ($t(185) = 4.336$, $p < .001$). A further analysis of p-value shows that there were no statistically significance differences ($p > 0.5$) in two storage tools namely; external hard drives ($t(185) = 1.832$, $p = .069$) and flash disks ($t(185) = 1.832$, $p = .069$).

5.3.3.5. Data backup strategies

Researchers were asked if they had put in place some mechanism to protect their data from losses. Findings revealed that 178 (93.7%) said they made backups for their research data. A follow-up questionnaire required respondents to indicate the strategies they used to back up their data. To answer this question, means, standard deviations and independence t-tests were performed. Findings are presented in Table 29.

Table 29. Data backup strategies (N=187)

Storage facility	University	N	*n* & (%)	M	SD	t- value	Sig.
Copies are uploaded on Google Drive	UNI1	103	68 (66.0)	1.3398	.47596	5.518	.000*
	UNI2	84	81 (96.4)	1.0357	.18669		
Copies are uploaded on Drop Box	UNI1	103	68 (66.0)	1.3398	.47596	3.857	.000*
	UNI2	84	75(89.3)	1.1071	.31115		
Copies are kept in my email	UNI1	103	90 (87.4)	1.1262	.33371	1.232	.217**
	UNI2	84	78(92.9%)	1.0714	.25909		
Datasets are saved on external hard drives	UNI1	103	93 (90.3)	1.0971	.29752	1.644	.102**
	UNI2	84	81 (96.4%)	1.0357	.18669		
Copies of datasets are saved on a local server	UNI1	103	8 (7.8)	1.9223	.26896	-2.645	.009*
	UNI2	84	0(0)	2.0000	.00000		
Copies of data sets are saved on a central campus server	UNI1	103	9 (8.7%)	1.9126	.28377	-2.821	.005*
	UNI2	84	0(0)	2.0000	.00000		
Copies of datasets are saved on a web-based server	UNI1	103	14 (13.6%)	1.8641	.34438	-2.395	.018*
	UNI2	84	3 (3.6)	1.9643	.18669		
Copies of datasets are stored in a data repository	UNI1	103	7 (6.8)	1.9320	.25291	-2.462	.015*
	UNI2	84	0(0)	2.0000	.00000		
Backup files are automatically generated	UNI1	103	0 (0)	2.0000	.00000		a
	UNI2	84	0 (0)	2.0000	.00000		
Backup files are manually generated	UNI1	103	7 (6.8)	1.9320	.25291	-2.462	.015*
	UNI2	84	0(0)	2.0000	.00000		

Note. [a.] t was not computed because the standard deviations of both groups were 0, *($p \leq 0.5$), **($p > 0.5$), constant interval percentage = 95%

Findings presented in Table 29 show that 68 (66%) UNI1 researchers used Google Drive, 68 (66%) used DropBox, 90 (87.4%) used email accounts and 93 (90.3) used external hard drives. Similarly, 81 (96.4) UNI2 researchers used Google Drive, 75 (89.3%) used DropBox and 78 (92.9%) used external hard drives. Results show further that nearly all researchers did not use various storage tools such as local servers, central campus servers, web-based servers, data repositories and automatic generation of servers.

Means, standard deviations and independent t-tests were computed to explore if there were statistically significant differences among researchers by university affiliations. Analysis of means and standard deviations reveal that overall, there were statistically significant differences across all the 10 dimensions with lowest mean (*M* = .47596) and highest mean (*M*

= 2.0000) while lowest standard deviation was (*SD* = .00000) and highest standard deviation was (*SD* = .47596). Analysis of the *t*-value and *p*-value reveal that specifically, statistically significance differences ($p \leq 0.5$) were observed in Google Drive ($t(185) = 5.518$, $p < .001$); DropBox ($t(185) = 3.857$; $p < .001$); local servers ($t(185) = -2.645$, $p = .009$); central campus server ($t(185) = -2.821$, $p = .005$); web-based severs ($t(185) = -2.395$, $p = .018$); data repositories ($t(185) = -2.462$, $p = .015$); and automatic generation of backups ($t(185) = -2.462$, $p = .015$). On the other hand, no statistically significant differences ($p > 0.5$) were noted in emails ($t(185) = 1.232$, $p = .217$) and in external hard drives ($t(185) = 1.644$, $p = .102$).

5.3.3.6. Infrastructure to support research data management

In this item, researchers were asked to indicate if their universities offered enough infrastructure to support management of research data they generated. Only 51 (26.8) researchers said yes while 136 (71.6%) said no. A follow-up question required respondents to indicate the kind of support that they wanted their universities to provide in order to strengthen research data management activities. The question was in form of a Lickert scale and researchers were required to indicate the extent to which they wanted their universities to provide each kind of support by choosing from the options of *Agree Strongly*, *Agree Somewhat*, *Neutral*, *Disagree Somewhat* and *Disagree Strongly*. Findings are presented in Table 30. An independent t-test was also computed to explore differences in kind of support needed by university affiliation.

Table 30. Support to be provided by universities (N=187)

Factors	Institution									
	UNI1 (***n*=103**)					UNI2 (***n*=84**)				
	Agree Strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Establish a process for managing data 5 years or less	79 (76.7%)	12 (11.7%)	0(%)	12 11.7%	0(0%)	58 (69.0%)	6 (7.1%)	20 (23.8%)	0(%)	0(%)
Establish a process for managing data beyond 5 years	72 (69.9%)	14 (13.6%)	1 (1%)	4 (3.9%)	12 (11.7%)	49 (58.3%)	6 (7.1%)	23 (27.4%)	0(%)	6 7.1%
Establish technical support for data management	63 (61.2%)	28 (27.2%)	12 11.7%	0(0%)	0(0%)	23 (27.4%)	3 (3.6%)	38 (45.2%)	20 (23.8%)	0(%)
Establish funds to support data management	58 (56.3%)	33 (32.0%)	12 (11.7%)	0(0%)	0(0%)	20 (23.8%)	9 (10.7%)	0(0%)	20 (23.8%)	35 (41.7%)

Independent t-test results

[1]($t(185) = -2.068$, $p = .040$); [2]($t(185) = -3.037$, $p = .003$); [3]($t(185) = -1.541$, $p = .125$); [4]($t(185) = -4.999$, $p < 001$).

Note. Constant interval percentage = 95%

Data presented in Table 30 show that 79 (76.7%) UNI1 researchers agreed strongly, 12 (11.7%) agreed somewhat and 12 (11.7%) were neutral on the need to establish a process for managing data for a period of five years or less; 72 (69.9%) UNI1 researchers agreed strongly, 14 (13.6%) agreed somewhat, one (1%) was neutral, four (3.9%) disagreed somewhat and 12 (11.7%) disagreed strongly on the need to establish a process for managing data beyond a period of five years; 63 (61.2%) agreed strongly, 28 (27.2%) agreed somewhat and 12 (11.7%) were neutral on the need to establish technical support for data; and 58 (56.3%) agreed strongly, 33 (32%) agreed somewhat and 12 (11.7%) were neutral on the need to establish funds to support research data management. For UNI2 researchers, 58 (69%) agreed strongly, six (7.1%) agreed somewhat and 20 (23.8%) were neutral on the need to establish a process for managing data for a period of five years or less; 23 (27.4%) agreed strongly, three (3.6%) agreed somewhat, 38 (45.2%) were neutral, 20 (23.8%) disagreed strongly on the need to establish funds to support research data management, and 49 (58.3%) agreed strongly, six (7.1%) agreed somewhat, 23 (27.4%) were neutral and six (7.1%) disagreed strongly on the need to establish a process for managing data beyond a period of five years.

Analysis of these results revealed that whereas UNI1 researchers agreed with the need for universities to provide all four types of research data management support services, UNI2 researchers only agreed with two of them which included the need to establish a process for managing data five years or less and establishing a process for managing data beyond five years.

These results are in support as well as in contrast with those realised from library staff on this theme. As can be observed in section 5.2.3.6, like researchers at UNI1, library staff were also of the view that the university should establish a process for managing data for a period of five years and beyond, establish technical support for data management and establish funds for supporting data management. However, while library staff at UNI2 were of the view that their university should establish technical support for data management only, researchers at UNI2 were of a different view that their college needed to establish a process for managing data for a period of five years and beyond.

An independent t-test was computed to determine if there were statistically significant differences between UNI1 and UNI2 in terms of the support they required their individual universes to provide. Results show that there were statistically significant differences in three dimensions ($p \leq 0.5$) and they included establishing a process for managing data for a period of five years or less ($t(185) = -2.068$, $p = .040$); establishing a process for managing data beyond a period of five years ($t(185) = -3.037$, $p = .003$); and establishing funds to support data management ($t(185) = -4.999$, $p < 001$). On the other hand, there were no statistically significant differences ($p > 0.5$) in the dimension of establishing funds to support data management ($t(185) = -1.541$, $p = .125$).

5.3.4. Competencies required for research data management

This section of questionnaire (See Appendix B) investigated of the various aspects about competencies required by researchers to proficiently partake RDM activities. Fours concepts were investigated and they included training workshops or training, metadata types, competencies in various RDM activities and the extent to which researchers sought support from various professionals.

5.3.4.1. Training workshops in research data management

This question set out to learn from researchers if they had attended any workshops or trainings related to research data management. Only a small number of 46 (24.2%) researchers indicated that they had ever attended workshops on research data management while the majority with scores of 141 (74.2%) said they had not attended any. For individual universities, 23 (22.3%) UNI1 researchers said they had attended and 23 (27.4%) UNI2 researchers said they had attended. For those researchers who had attended the workshops, they were asked to mention organisers of those workshops by indicating if the workshops were organised by their university, their university library, a local organisation or international organisation. Findings reveal that of the 46 researchers who attended such workshops, 35 (76.1%) said they were organised by international organisations. The other 11 (23.9%) researchers did not indicate the organisers of the workshops or training.

5.3.4.2. Metadata used by researchers

Two questionnaire items gathered data on metadata. First, researchers were asked to indicate if they assigned metadata to their research data. With the two universities combined, only 14 (7.4%) said they assigned some metadata to their research data while the majority with a score

of 173(91.1%) said they did not. For individual universities, five (4.9%) and nine (10.7%) at UNI1 and UNI2 respectively said they assigned metadata. Second, researchers who said that they assigned metadata were provided with a list of types of metadata to indicate the ones they assigned to the research data they generated and preserved. The list of metadata types included metadata standardised within laboratories, International Standards Organisation (ISO), Open GIS, Ecological Metadata Language, Federal Geographic Data Committee, Dublin Core, Darwin Core and Directory Interchange Format. Findings revealed that nine (4.7%) researchers used metadata within their laboratories, two (1%) used Dublin Core, another two (1%) used Ecological Metadata Language and only one (.5%) used Federal Geographic Data Committee. The results show that most researchers at both universities do not assign metadata to the research data they generated and preserved.

5.3.4.3. Competencies in research data management

The aim of this item was to find out the competency of researchers in performing specific research data management activities. To achieve the aim of this item, the researcher listed some activities involved in managing research data and researchers were asked to indicate if they were competent or if they needed to be trained. This item was investigated further by performing an independent t-test. Findings are presented in Table 31.

Table 31. Skills for researchers in managing research data (N=187)

Dimension	University	N	Competent***	Need training***	M	SD	t-value	Sig.
Writing data management plans	UNI1	103	33 (32.0%)	70(68%)	1.6796	.46891	-2.872	.005*
	UNI2	84	12 (14.3%)	72(85.7%)	1.8571	.35203		
	Total	187	45 (24.1 %)	142(75.9%)	-	-		
Advanced computing	UNI1	103	26 (25.2%)	77(74.8%)	1.7476	.43653	-2.564	.011*
	UNI2	84	9 (10.7%)	75(89.3%)	1.8929	.31115		
	Total	187	35 (18.7%)	152(81.3%)	-	-		
Short term data preservation strategies	UNI1	103	9(8.7%)	94(91.3%)	1.9126	.28377	-.475	.635**
	UNI2	84	12(14.3%)	72 (85.7)	1.9643	1.05792		
	Total	187	21(11.2%)	166(88.2%)	-	-		
Long term data preservation strategies	UNI1	103	9(8.7%)	94(91.3%)	1.9126	.28377	-.397	.691**
	UNI2	84	6 (7.1%)	78(92.9%)	1.9286	.25909		
	Total	187	15(8.0%)	172(92.0%)	-	-		
Preservation planning	UNI1	103	22(21.4%)	81(78.6%)	1.7864	.41185	-1.712	.089**
	UNI2	84	10(11.9%)	74(88.1%)	1.8810	.32579		
	Total	187	32(17.1%)	155(82.9%)	-	-		
Identifying standards and practices	UNI1	103	13(12.6%)	90(87.4%)	1.8738	.33371	-2.218	.028*
	UNI2	84	3(3.6%)	81(96.4%)	1.9643	.18669		
	Total	187	16(8.6%)	171(91.4%)	-	-		
Creating preservation metadata	UNI1	103	5(6.0%)	79(94%)	1.9515	.21596	.330	.742**
	UNI2	84	6 (7.1%)	78(92.9%)	1.9405	.23802		
	Total	187	10(5.3%)	177(94.7%)	-	-		
Depositing data into repositories or archives	UNI1	103	4(3.9%)	99(96.1%)	1.9612	.19415	.983	.327**
	UNI2	84	6(7.1%)	78(92.9%)	1.9286	.25909		
	Total	187	10(5.3%)	177(94.7%)	-	-		
Adhering to data management standards	UNI1	103	4(3.9%)	99(96.1%)	1.9223	.26896	.983	.227**
	UNI2	84	6(7.1%)	78(92.9%)	1.9643	.18669		
	Total	187	10(5.3%)	177(94.7%)	-	-		
Disposition of data	UNI1	103	5(4.9%)	98(95.1%)	1.9515	.21596	1.516	.131**
	UNI2	84	9(10.7%)	75(89.3%)	1.8929	.31115		
	Total	187	14(7.5%)	173(92.5%)	-	-		
Migrating data to newer file formats	UNI1	103	6(5.8%)	97(94.2%)	1.9417	.23537	-.713	.477**
	UNI2	84	3(3.6%)	81(96.4%)	1.9643	.18669		
	Total	187	9(4.8%)	178(95.2%)	-	-		

Note. *($p \leq 0.5$), **($p > 0.5$); ***percentages calculated against total number of individual universities i.e. % of 103 for UNI1 and % of 87 for UNI2; percentages in the *total row* are calculated against the total number i.e. % of 187, constant interval percentage = 95%.

Generally, findings presented in Table 31 show that researchers at both universities lack various research data management skills hence the need for them to be trained. For UNI1 researchers, 72 (85.7%) needed training in writing data management plans; 77 (74.8%) needed training in advanced computing; 94 (91.3%) needed training in short and long term data preservation strategies; 81(78.6%) needed training in preservation planning; 90 (87.4%) needed training in identifying new standards, practices and software for curation; 79 (94%) needed training in creating preservation metadata for describing data sets; 99 (96.1%) needed training in depositing data into repositories or archives; 99 (96.1%) needed training in storing

digital information in a secure manner adhering to relevant standards; 98 (95.1%) needed training disposing data not selected for long term preservation; and 97 (94.2%) needed training in migrating digital information to newer file formats that support its continued access and preservation. Likewise, Table 31 shows that for UNI2 researchers, 70 (68%) needed training in writing data management plans; 75 (89.3%) needed training in advanced computing; 72 (85.7%) needed training in short term data preservation strategies; 78 (92.9%) needed training in long-term data preservation strategies; 74 (88.1%) needed training in preservation planning; 81(96.4%) needed training in identifying new standards, practices and software for curation; 78 (92.9%) needed training in creating preservation metadata for describing data sets; 78 (92.9%) needed training in depositing data into repositories or archives; 78 (92.9%) needed training in storing digital information in a secure manner adhering to relevant standards; 75 (89.3%) needed training in disposing of data not selected for long-term preservation; and 81(96.4%) needed training in migrating digital information to newer file formats that support its continued access and preservation.

Data on this item was analysed further using the independent t-test to determine if there were statistically significant differences in specific researchers' competencies in relation to university affiliation. Means and standard deviations show that generally, there were no statistically significant differences as it can be seen in Table 31 where the lowest mean was (M= 1.6796) and the highest mean was (M=1.9643) and the minimum standard deviation was (SD = .18669) and the maximum was (M = 1.05792). Analysis of the p-value also reveals that most or eight dimensions showed no statistically significant differences ($p > 0.5$) and they include short term data preservation strategies ($t(185)$ = -.475, p = .635); long term data preservation strategies ($t(185)$ = -.397, p = .691); preservation planning ($t(185)$ = -1.712, p = .089); creating preservation metadata ($t(185)$ = .330; p = .742); depositing data into repositories or archives ($t(185)$ = .983, p = .327); adhering to data management standards ($t(185)$ = .983, p = .227); disposition of data not selected for preservation ($t(185)$ = 1.516, p = .131); and migrating data to newer file formats ($t(185)$ = -.713; p = .477). Analysis of the p-value reveals further that statically significant differences ($p \leq 0.5$) between UNI1 and UNI2 researchers occurred in three dimensions which include writing data management plans ($t(185)$ = -2.872, p = .005); advanced computing ($t(185)$ = -2.564, p = .011) and identifying standards and practices ($t(185)$ = -2.218, p = .028).

5.3.4.4. Frequency with which researchers seek support from some professionals

In this item, researchers were asked to indicate how frequently they sought help from various professional, vis-à-vis librarians, ICT experts, directors of research and fellow researchers. Findings show that in terms of seeking help from ICT experts, one (.5%) said always, 54 (28.9%) said frequently, 72 (38.5%) said occasionally, 11 (5.9%) were not sure and 49 (26.2%) said never. In terms of help from librarians, two (1.1%) said always, 77 (41.2%) said frequently, 99 (52.9%) said occasionally and nine (4.8%) said never. Thirty two (17.1%) said they always sought help from fellow researchers, 63 (33.7) said frequently, 33 (17.6%) said occasionally, eight were not sure and 51 (27.3%) said never. On the frequency they sought help from directors of research, only one (.5%) said always while 186 (99.5%) said never. These results suggest that researchers mostly seek help from librarians in their research data management activities. These results give credence to those reported in section 5.2.2.1 where it was noted that the majority of library staff with scores of 27 (75%) indicated that researchers consulted them on research activities.

5.3.5. Challenges in research data management

This section of the questionnaire (See Appendix B) had two items that aimed at identifying challenges related to research data management amongst researchers. Specific issues that were investigated included data loss amongst researchers and the challenges that those researchers faced in managing their data.

5.3.5.1. Data loss amongst researchers

This item sought to find out from researchers the frequency of losing digital research data based on three factors that included stolen storage facilities, accidental damage of storage facilities and obsolescence of technologies. For each factor, researchers were asked to select amongst frequently, occasionally, not sure and never. An independent t-test was performed to find out if there were differences between the two universities in data loss through each factor. Findings are presented in Table 32.

Table 32. Data loss amongst researchers (N=187)

Dimension	University	N	Frequently**	Occasionally**	Not sure **	Never**	M	SD	t-value	Sig.
Stolen storage facilities	UNI1	103	0 (0%)	20 (19.4%)	32 (31.1%)	51 (49.5%)	3.6990	.77756	-4.196	.000*
	UNI2	84	3 (3.6%)	6 (7.1%)	0 (0%)	75 (89.3%)	3.1071	.56007		
	Total	187	3 (1.6%)	126 (67.4%)	32 (17.1%)	26 (13.9%)	-	-		
Accidental damage	UNI1	103	45 (43.7%)	54 (52.4%)	0 (0%)	4 (3.9%)	2.6311	.70000	-4.648	.000*
	UNI2	84	3 (3.6%)	78 (92.9%)	0 (0%)	3 (3.6%)	3.0357	.42359		
	Total	187	48 (25.7%)	132 (70.6%)	0 (0%)	7 (3.7%)	-	-		
Obsolescence of technologies	UNI1	103	6 (5.8%)	7 (6.8%)	13 (12.6%)	77 (74.8%)	4.5631	.85943	5.847	.000*
	UNI2	84	0 (0%)	0 (0%)	3 (3.6%)	81 (96.4%)	4.9643	.18669		
	Total	187	6 (3.2%)	7 (3.7%)	16 (8.6%)	158 (84.5%)	-	-		

Note. *($p \leq 0.5$); **percentages calculated against total number of individual universities i.e. % of 103 for UNI1 and % of 87 for UNI2; percentages in the *total row* are calculated against the total number i.e. % of 187, constant interval percentage = 95%.

In terms of stolen storage facilities, Table 22 shows that 20 (19.4%) UNI1 researchers indicated occasionally, 32 (31.1%) were not sure and 51(49.5%) said never. Three (3.6%) UNI2 staff said frequently, six (7.1%) said occasionally and 75 (89.3%) said never. On the aspect of accidental damage, 45(43.7%) UNI1 researchers indicated frequently, 54 (52.4%) said occasionally and four (3.9%) said never. Three (3.6%) UNI2 researchers said frequently, 78 (92.9%) indicated occasionally and three (3.6%) said never. Finally, on the aspect of obsolescence of technologies, six (5.8%) UNI1 researchers said frequently, seven (6.8%) indicated occasionally, 13 (12.6%) were not sure; and 77 (74.8%) said never. Three (3.6%) UNI2 researchers were not sure and 81(96.4%) said never.

To determine if there were statistically significant differences in research data loss by university affiliation, means, standard deviations and independent t-tests were computed. There were generally significant differences across the three dimensions as means ranged from (M = 2.6311) to (M = 4.9643) and standard deviations ranged from (SD = .18669) to (SD = .85943). General distributions of means and standard deviations are supported by the outcomes of the analysis of the t-value and p-value which shows that there were statistically significant differences across the three dimensions; stolen storage facilities ($t(185)$ = -4.196, p < .001*);*

accidental damage ($t(185)$ = -4.648, p < .001); and obsolescence of technologies ($t(185)$ = 5.847, p < .001).

5.3.5.2. Challenges researchers face in managing research data

The respondents were presented with a list of factors that could affect their research data management activities. For each factor, researchers were asked to indicate the extent to which each of the factors affected their involvement in research data management by selecting one option from the following: agree strongly, agree somewhat, neutral, disagree, somewhat and disagree strongly. Results are captured in Table 33. ANOVA was also computed to explore differences in challenges affecting research data management between the two universities.

Table 33. Challenges researchers face in managing research data (N=187)

Factors	Institution									
	UNI1 (*n*=103)					UNI2 (*n*=84)				
	Agree Strongly	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Lack of incentives to share data[1]	58 (56.3%)	31 30.1%	0(%)	1 (1%)	13 (12.6%)	0(%)	35 (41.7%)	14 (16.7%)	23 (27.4%)	12 (14.3%)
Lack of storage and network infrastructure[2]	29 (28.2%)	53 (51.5%)	19 (18.4%)	2 (1.9%)	0(%)	35 (41.7%)	0(%)	6 (7.1%)	23 (27.4%)	20 (23.8%)
Lack of curation tools and software[3]	15 (14.6%)	25 (24.3%)	20 (19.4%)	16 (15.5%)	(27 26.2%)	47 (56.0%)	28 (33.3%)	9 (10.7%)	0(%)	0(%)
Lack of policy frameworks[4]	58 (56.3%)	34 (33.0%)	0(%)	9 (8.7%)	0(%)	0(%)	35 (41.7%)	14 (16.7%)	23 (27.4%)	12 (14.3%)
Lack of curation skills and training[5]	53 (51.5%)	42 (40.8%)	0(%)	8 (7.8%)	0(%)	49 (58.3%)	23 (27.4%)	0(%)	9 (10.7%)	3 (3.6%)
Lack of guidance and support[6]	47 (45.6%)	48 (46.6%)	0(%)	8 (7.8%)	0(%)	49 (58.3%)	23 (27.4%)	0(%)	9 (10.7%)	3 (3.6%)
Difficulty in finding data produced by others[7]	43 (41.7%)	54 (52.4%)	5 (4.9%)	0(%)	1 (1.0%)	21 (25.0%)	57 (67.9%)	5 (6.0%)	0(%)	1 (1.2%)
Most data is not trustworthy[8]	4 (3.9%)	7 (6.8%)	8 (7.8%)	36 (35.0%)	48 (46.6%)	4 (4.8%)	1 (1.2%)	9 (10.7%)	53 (63.1%)	17 (20.2%)
Lack of skills in sharing data[9]	52 (50.5%)	43 (41.7%)	0(%)	8 (7.8%)	0(%)	49 (58.3%)	23 (27.4%)	0(%)	9 (10.7%)	3 (3.6%)
Tracking updates to data [10]	58 (56.3%)	45 (43.7%)	0(%)	0(%)	0(%)	58 (69.0%)	23 (27.4%)	0(%)	0(%)	3 (3.6%)
Lack of skills to create metadata[11]	55 (53.4%)	46 (44.7%)	2 (1.9%)	0(%)	0(%)	58 (69.0%)	23 (27.4%)	0(%)	0(%)	3 (3.6%)
Lack of standardised metadata[12]	0(%)	23 (22.3%)	80 (77.7%)	0(%)	0(%)	0(%)	20 (23.8%)	64 (76.2%)	0(%)	0(%)
Failure by data re-users to cite my the data[13]	4 (3.9%)	13 (12.6%)	0(%)	41 (39.8%)	45 (43.7%)	0(%)	0(%)	9 10.7%	52 (61.9%)	23 (27.4%)
Lack of support from the university[14]	64 (62.1%)	34 (33.0%)	1 (1.0%)	3 (2.9%)	1 (1.0%)	0(%)	49 (58.3%)	9 (10.7%)	20 (23.8%)	6 (7.1%)
Prohibitive institutional policies[15]	4 (3.9%)	7 (6.8%)	8 (7.8%)	36 (35.0%)	48 (46.6%)	0(%)	0(%)	9 (10.7%)	58 (69.0%)	17 (20.2%)
Obsolescence of technologies[16]	2 (1.9%)	0(%)	0(%)	31 (30.1%)	70 (68.0%)	0(%)	0(%)	1 (1.2%)	27 (32.1%)	56 (66.7%)
Ethical and legal norms[17]	5 (4.9%)	31 (30.1%)	47 (45.6%)	19 (18.4%)	1 (1.0%)	58 (69.0%)	12 (14.3%)	0(%)	14 (16.7%)	0(%)

ANOVA

[1]($F(1,185) = 3.691$, $p = .056$); [2]($F(1,185) = 1.332$, $p = .250$); [3]($F(1,185) = -88.989$, $p < .001$); [4]($F(1,185) = -3.494$, $p = .063$); [5]($F(1,185) = .456$, $p = .501$); [6]($F(1,185) = .074$, $p = .786$); [7]($F(1,185) = 3.753$, $p = .054$); 8($F(1,185) = 2.004$, $p = .159$); [9]($F(1,185) = .370$, $p = .544$); [10]($F(1,185) = .043$, $p = .836$); [11]($F(1,185) = .470$, $p = .494$); [12]($F(1,185) = .057$, $p = .812$); [13]($F(1,185) = .514$, $p = .474$); [14]($F(1,185) = 101.575$, $p < .001$); [15]($F(1,185) = .099$, $p = .754$); [16]($F(1,185) = .138$, $p = .711$). [17]($F(1,185) = 66.762$, $p < .001$).

Note. Constant interval percentage = 95%

Findings presented in Table 33 show that while some factors affected researchers' involvement in research data management activities, others did not. Specifically, 58 (56.3%) UNI1 staff agreed strongly, 31 (30.1%) agreed somewhat, one (1%) disagreed somewhat and 13 (12.6%) disagreed strongly with the factor of lack of incentives to share data. Pertaining to lack of storage and network infrastructure, 29 (28.2%) agreed strongly, 53 (51.5%) agreed somewhat, 19 (18.4%) were neutral and two (1.9%) disagreed somewhat. In terms of lack of policy frameworks, 58 (56.3%) agreed strongly, 34 (33%) agreed somewhat and nine (8.7%) disagreed somewhat. Fifty three (51.5%) agreed strongly, 42 (40.8%) agreed somewhat and eight (7.8%) disagreed somewhat with the factor of lack curation skills and training; 47 (45.6%) agreed strongly, 48 (46.6%) agreed somewhat, and eight (7.8%) disagreed somewhat on lack of guidance and support; 43 (41.7%) agreed strongly, 54 (52.4%) agreed somewhat, five (4.9%) were neutral, one (1%) disagreed strongly that it was difficult in finding data produced by others; 52 (50.5%) agreed strongly, and 43 (41.7%) agreed somewhat and eight (7.8%) disagreed somewhat about lack of skills in sharing data; 58 (56.3%) agreed strongly and 45 (43.7%) agreed somewhat on the factor of tracking updates to data; 55 (53.4%) agreed strongly, 46 (44.7%) agreed somewhat and two (1.9%) were neutral on the aspect of lack of skills to create metadata; and finally, 64 (62.1%) agreed strongly, 34 (33%) agreed somewhat, one (1%) was neutral, three (2.9%) disagreed somewhat and one (1%) disagreed strongly on the aspect of lack of support from the university.

For UNI2, results presented in Table 33 show a number of factors that affected UNI2 researchers in their research data management activities. Specifically, 47 (56%) agreed strongly, 28 (33.3%) agreed somewhat and nine (10.7%) were neutral on lack of curation tools and software; 49 (58.3%) agreed strongly, 23 (27.4%) agreed somewhat, nine (10.7%) disagreed somewhat and three (3.6%) disagreed strongly about lack of curation skills and training; 49 (58.3%) agreed strongly, 23 (27.4%) agreed somewhat, nine (10.7%) disagreed somewhat and three (3.6%) disagreed strongly about lack of guidance and support; and 21 (25%) agreed strongly, 57 (67.9%) agreed somewhat, five (6%) were neutral and one (1%) disagreed strongly that it was difficult to find data produced by other researchers. Results in Table 33 shows further that 49 (58.3%) UNI2 researchers agreed strongly, 23 (27.4%) agreed somewhat, nine (10.7%) disagreed somewhat and three (3.6%) disagreed strongly about lack of skills in sharing data; 58 (69%) agreed strongly, 23 (27.4%) agreed somewhat and three (3.6%) disagreed strongly about the inability to track updates to data; and finally, 58 (69%)

agreed strongly, 23 (27.4%) agreed somewhat and three (3.6%) disagreed strongly that they lacked skills in creating metadata.

From these results, it is clear that the common factors that affected researchers at both universities included the lack of curation skills and training, lack of guidance and support, difficulty in finding data produced by other researchers, lack of skills in sharing data, inability to track updates to data and lack of skills to create metadata. Factors that applied to UNI1 researchers only included lack of incentives to share data, lack of storage and network infrastructure and lack of support from the university. Lack of curation tools and software is a factor that applied to UNI2 researchers only. However, researchers from both universities did not agree with some factors that included that most data was not trustworthy, lack of standardised metadata, failure by data re-users to cite researchers' data, prohibitive institutional policies, obsolescence of technologies and ethical and legal norms.

ANOVA was computed in order to determine if there were statistically significant differences in the factors that affected researchers of the two universities in their research activities. Results revealed that there were no statistically significant differences in 13 factors namely, lack of incentives to share data ($F(1,185) = 3.691$, $p = .056$); lack of storage and network infrastructure ($F(1,185) = 1.332$, $p = .250$); lack of policy frameworks ($F(1,185) = -3.494$, $p = .063$); lack of curation skills and training ($F(1,185) = .456$, $p = .501$); lack of guidance and support ($F(1,185) = .074$, $p = .786$); data not trustworthy ($F(1,185) = 2.004$, $p = .159$); lack of skills in sharing data ($F(1,185) = .370$, $p = .544$); tracking updates to data ($F(1,185) = .043$, $p = .836$); lack of skills to create metadata ($F(1,185) = .470$, $p = .494$); lack of standardised metadata ($F(1,185) = .057$, $p = .812$); failure by data re-users *to* cite researchers' data ($F(1,185) = .514$, $p = .474$); prohibitive institutional policies ($F(1,185) = .099$, $p = .754$); obsolescence of technologies ($F(1,185) = .099$, P = .754). Analysis of the *f*-value and *p*-value showed further there were statistically significant differences in four dimensions that included lack of curation tools and software ($F(1,185) = -88.989$, $p < .001$); difficulty in finding data produced by others ($F(1,185) = 3.753$, $p = .054$); lack of support from the university ($F(1,185) = 101.575$, $p < .001$); and ethical and legal norms ($F(1,185) = 66.762$, $p < .001$).

5.4. Presentation of qualitative data from directors of research

As stated in Chapter Four, the study also collected qualitative data by conducting in-depth interviews with directors of research from each targeted university. An interview guide (See Appendix C) was used to guide the researcher in conducting the interview. This section

presents data on the four themes that are the cornerstone of the study namely, research data creation, sharing and re-use practices; research data preservation practices; competencies in data curation; and challenges that affect the management of research data. Considering that the study adopted the side-by-side analysis approach (see Chapter Four), in this section efforts are also made by the researcher to triangulate the findings by using qualitative data to support or contradict part of the findings realised from quantitative data which were presented in sections 5.2 and 5.3.

5.4.1. Demographic data

In terms of qualification, directors of research from both universities were holders of PhDs. They were both holding the rank of Associate Professor. They were both males and had worked at their respective universities for over 15 years in various capacities. As research directors, their key roles included coordinating all research activities taking place at the university and representing their universities in all research activities at national, regional and international levels. For example, Director of Research at UNI2 (DR- UNI2) said that "as a director of research, my duty is to oversee all research activities in the college and making sure that researchers are assisted in grant management, ethical clearance issues, you know, they [researchers] face hiccups and my office is there to make sure these things move smoothly". In terms of gender, these results are similar to those reported in sections 5.5.1.2 and 5.3.1.3 where it was found that there were more males than females implying that the university system in Malawi is dominated by males.

5.4.2. Research data creation, sharing and re-use practices

Directors of research were interviewed on various themes concerning research data creation, sharing and re-use. The aim of this section is to report key findings in relation to these themes. The section also reports on the importance of research, a theme that emerged unexpectedly in the course of conducting the interviews.

5.4.2.1. Importance of research

This theme emerged in the course of the interview process as it was not one of the items in the interview guide. Both directors of research decided to state the importance of research in their universities and from their personal perspectives. Findings revealed that both respondents indicated that research is important in various ways. Among others, both respondents said research drives national economies through the discovery of new knowledge, it helps to expose

universities to the international landscape, and it is a source of income for universities. From their personal perspective, research helps academics to rise in their academic career through promotions to various higher ranks within the university such as Senior Lecturer, Associate Professor and so on. In relation to this aspect, the following comments were made by the two directors of research:

> **DR-UNI2:** *In my professional life, research is extremely important not just important in the sense that in [an] academic environment, as academics, for you to progress you need to research. For the college, I think it's our number one earner of finances, so research is the one that keeps this college running.*

> **DR-UNI1:** *Basically, research is very important for* UNI1 *because it helps us on regional and international university ranking, and more importantly, it helps our academics to get promoted.*

5.4.2.2. Research output

During the interview, the researcher wanted to know from the respondents the strategies their universities have put in place to maximise or encourage researchers to conduct extensive research and publish outcomes in creditable journals. Findings revealed that both universities have put in place mechanisms to boost research publications amongst their researchers. The key mechanisms at both universities included rewarding researchers who published by promoting them to higher ranks and helping researchers pay article publication processing fees in open access journals and sponsoring researchers to attend and present research findings in national, regional and international conferences. However, while UNI2 was rewarding its researchers with USD100 for each publication in creditable journals, UNI1 had not yet started rewarding its researchers through this arrangement. Again, while UNI2 organised a Research Dissemination Conference every year, UNI1 did not. Findings showed that one common theme emerging from the study was that both institutions rarely sponsored their researchers to conduct studies. A selection of verbatim quotes from the interviews with directors of research is presented below.

> DR-UNI1**:** *So, in the meantime, the only way we encourage researchers to publish is through promotion, that is, when academics publish, they get promoted. Very soon however, we want to start rewarding them through monitory incentives whereby when a researcher publishes in a reputable journal, they should receive some money as part of encouraging them to publish more. Where researchers are supposed to pay*

processing fees and indeed, where they cannot manage to pay, the university comes to their rescue and supports payment of publication fees.

DR-UNI2: *At [the] university, level of course, there are policies that you know give flexibilities to every lecturer to be conducting research. At [the] college level, we make deliberate policies where people are allowed to collaborate with other people and affiliates so that they can be able to conduct research. And if they conduct research and publish, at the college level, we reward them. For each publication, we give them USD100 to keep their research going. So those are some of the things but people are encouraged at university and college level. More importantly, as I already said, when they publish, they get promoted to senior positions, we have the Annual Research Conference which is also one way we share the data/research findings at the college.*

The fact that UNI2 researchers received monetary incentives while UNI1 researchers did not may explain the reason why results reported in section 5.3.2.1 showed that generally, more UNI2 researchers had more papers published, in review, and commissioned reports than UNI1 researchers. The results may further support the findings reported in section 5.3.2 where it was noted that unlike researchers at UNI2, the majority of researchers at UNI1 did not share their data because of lack of incentives (See sections 5.3.2.6 and 5.3.4.2), lack of funding, and lack of support from university (See section 5.3.4.2).

5.4.2.3. The concept of digital research data

The researcher asked directors to comment on how they understood the concept of digital research data. Results revealed that respondents' explanation of the concept was sensible and was within the acceptable universal tenets that define digital research data or data curation. It was further noted that both respondents' explanations were influenced by flexibilities in sharing digital research data. Comments from respondents are presented below.

DR-UNI2: *…its new concept of course, but I understand it as where researchers from different institutions can share the data as well as you can re-use that data for future discoveries. It's not like the old time when you have your data and once you publish, you destroy it, nowadays, you keep it so that other people can come with another angle to look at that data and new discoveries can come out from that data. Because it is in digital form, it becomes much easier [to share].*

DR-UNI1: *From the directorate point of view and also as a researcher, it's a good thing because I know it can also be a resource, it's also money, If you are well organised and there are good polices, it's something that should be encouraged because it can be a source of revenue for the university. Basically, digital research data is in digital format, it is easy to share, preserve and re-use.*

5.4.2.4. Ownership of research data

The researcher asked respondents to explain issues surrounding intellectual property rights regarding data generated by its lecturers or researchers in their universities. The study found that issues of data ownership were treated completely differently at the two institutions. At UNI2, all the research data produced by its staff belonged to the college, this was regardless of whether the data was generated through self-sponsored research, and university funded research as well as donor funded research. According to the Director of Research at UNI2, this was the case because researchers were using the college name and facilities to conduct their research activities. On the other hand, at UNI1, data generated from self-sponsored research belonged to individual researchers and data generated through donor funded research belonged to the donors or their collaborators. According to the Director of Research at UNI1, the contributing factor was that the university did not have policies in place concerning research data management. As already noted, both institutions rarely funded researchers to conduct research though UNI2 researchers received funding from external research grants organisations. Below are some quotes that were extracted from the interview that the researcher conducted with respondents.

DR-UNI2: *Ok, so it's whenever you are doing research in the university, whether you are using your own money or donor money, the intellectual property rights are of the university because you are using the university name and facilities. So the university owns that data, the holder of the intellectual rights is the university.*

DR-UNI1*: So, maybe let's start by pointing out that at the moment, we do not have a clear policy and here we are talking about ownership of research data. If it's an independent study, the output of the research and data belong to the researcher and we also know that if it's funded by the donor, the same things apply - the donor will have a say over the research and the data and sometimes you publish in open access [journals or databases] so that it's accessible for free.*

These results are in line with the findings realised from researchers as reported in section 5.3.24 that while none of UNI1 researchers was compelled by university policy to share their data, some researchers at UNI2 were compelled by the college research policy. The findings imply that in the UNI2 research policy, there is an element of data ownership and sharing while UNI1 does not even have a research policy.

5.4.2.5. Data sharing

One of the areas addressed during the interview was to find out the mechanisms universities put in place to foster research data sharing amongst researchers. Findings show that at UNI2, there were two key initiatives that encourage researchers to share research data. First, UNI2 established a data centre which is under the research support centre where there were people fully employed to manage data activities. Apart from the data centre, other departments have established their own data repositories which researchers used to share data with their collaborators. In addition, UNI2 made sure that when its researchers collaborated with international organisations or researchers in research endeavours, the college advised them to insert a clause on data sharing and accessibility. At UNI1, the office of the Director of Research indicated that the university has not put in place any mechanisms for encouraging data sharing among its research stakeholders. The only point worth mentioning is that the university encouraged researchers intending to collaborate with international research organisations in research activities to ensure that when preparing contracts, they were required to include mechanisms of data sharing. Like at UNI2, UNI1 researchers have had bad experiences with international research collaborators. International researchers come to collaborate with them but end up getting all the data and denying them access to the data such that local collaborating researchers cannot even publish from such research studies. Some of the verbatim responses from the interviews follows:

> **DR-UNI2:** *So, mostly in the university there is not really a problem, people are encouraged to share data because we have the data centre at the research support centre, there are people who are employed to specifically process the data so that people can have access to it. Nowadays, even researchers are encouraged - it's no longer that this is my data, you share with other people. But when it comes to international organisations, then we make sure that when signing contracts, the issue of data sharing should be clearly stipulated because sometimes people come here, they collect data and they go away with all of it while they were collaborating with us so we*

really make sure a clause on data sharing and access is included. We know that other projects [in the college] have their own databases so we have some units that are also processing data and maybe they do share with their collaborators. But we needed one central place where data is kept and in most universities it's actually the library which is the custodian of the data centres.

DR-UNI1*: To be honest even at the university level, I think, I doubt if there is any sharing of data collected by one particular study in the same department, faculty and even university. I really doubt but the reason is that we do not have a mechanism where after collecting data, you can deposit so that other people can access and use that data under certain terms. So the point I am making is that as a university, we do not have a mechanism that number one, here is the data, you should know that this data already exists, then this is how you can access and this should be true for our students as well.*

These results suggest that UNI2 has made some initiatives towards data sharing by establishing a data centre. On the other hand, UNI1 has not established any initiatives to encourage its researchers to share data. These findings show further that both universities play an important role in structuring contractual agreements with foreign collaborators by advising their researchers to include a clause on accessibility to data resulting from those collaborative research projects.

5.4.2.6. Data re-use

During the interview, respondents were asked to explain the initiatives their universities had taken to encourage researchers to use data produced by other researchers or research institutions. Findings revealed that at UNI2, researchers especially postgraduate students were being encouraged to re-use research data. The Director of Research at UNI2 did not mention any formal mechanisms employed by the college to encourage researchers to re-use data but focused more on providing reasons why the college encouraged researchers to re-use data. The reasons included the availability of large amounts of data laying idle at UNI2 and at the national level; less cost of re-using data than collecting fresh one; convenience on the part of re-users as there were no hurdles encountered in ethical clearance; and that re-use or re-analysis of previously generated data contributes to new knowledge breakthroughs. At UNI1, the university indicated that there were no formal procedures or mechanisms put in place to encourage researchers to re-use data mainly because there were no policies to guide such initiatives. The university was, however, in support of the concept of data re-use because of

three key reasons that were highlighted by the Director of Research. First, the university believed that combining data collected by different researchers or research groups could lead to landmark discoveries and secondly, re-using data is less costly than collecting new data; and finally, the university collected large amounts of data which was unused. A selection of quotes from the interview in relation to data re-use is presented below.

> **DR-UNI2:** *We really encourage the re-use of data because as a country, we have lots and lots of data which is just lying there and we encourage people especially postgraduate and undergraduate students to use it. You know conducting fresh research is costly but students at that level, at masters and undergraduate, they can just go into database and re-use that data and that becomes much easier - it's less costly and there are less hassles in ethical clearance. Yes, for my own students, since I do collect data, I usually have interns who come for maybe three months, so they really use my data. When patients come, we do collect their details and keep data in our databases and when students come, we tell them to use such data that is, compute correlations of certain variables for example.*
>
> **DR-UNI1***: Yes, we collect a lot of data but not all of it is used. You see another thing is you can collect data and another person collects data but you know when you combine the two sets, you can come up with a rich discovery. So to me, it's a matter of coming up with procedures that if you want to use this data, you have to acknowledge and I am very sure members of staff will start re-using data. But people must appreciate that it's cheaper to use data already generated than start afresh collecting the same data.*

The findings suggest that UNI2 was at least encouraging its researchers, particularly postgraduate students to re-use data though on small scale. On the one hand, UNI1 did not encourage its researchers to re-use data. It is therefore not surprising that it was found in section 5.3.2.8 that only 45 (23.7%) researchers said that they always used research data produced by others while the majority of researchers said they never used or used infrequently such data. Claims by the two directors of research further confirm results reported in section 5.3.3.3 that the majority of researchers produced data in capacities of one GB to one terabyte. It has to be acknowledged that although the researcher did not investigate the benefits of data sharing from the perspectives of library staff and researchers, this theme emerged unexpectedly during the interview with the two directors of research. The study revealed three common benefits

associated with data sharing and re-use namely; advancement of science through re-analysis of previously produced data; minimised costs for data re-use and convenience in research ethics clearance.

5.4.3. Research data preservation

Since both directors of research had indicated that the two universities generated large amounts of research data, the researcher wanted to know if their universities had adequate data storage infrastructure. For UNI2, considering that it had established a data centre, the question intended to know if there were other data storage facilities apart from the data centre and it was in the process of answering this question that the director revealed that the data centre itself was not adequate and sustainable. Results showed that not all researchers made use of the data centre because of two reasons. First, the data centre was offering services on a small scale due to limited capacity in terms of infrastructure and personnel. Technically, there was no centralised system for managing data at UNI2. Most data was kept in departmental laboratories in flash disks, external hard drives, computers or laptops and according to the Director of Research, the unfortunate part was that there were no proper back-ups. As a result, most researchers lost their data when computers or laptops crashed or were lost. Results showed that there were some sections within the university that had offshore data stores but the college's plan was to have a centralised data management centre where all data produced by its researchers or partners should be kept, controlled and managed. Second, most researchers at UNI2 were not aware of the research services offered at the data centre. The Director of Research at UNI2 emphasised that the college could not blame researchers for not making use of the facility because the college had not done enough to publicise its services to those researchers. The following is a verbatim response from the Director of Research at UNI2:

> *So the way we are doing [managing data] now is at a very small scale and it's not a sustainable way because researchers keep most of their research data in their laptops and computers and other storage facilities and once these crash, the data is lost. There is no central place where we have really dependable servers. The data is in the labs in the computers - it's only the research support centre and some departments that have their own databases but do not have proper backup systems. But we need to really buy our own servers with proper backup systems with an IT [Information Technology] person responsible for data so that whatever you are doing in the laboratory, you automatically send to the centralised data centre so that if you lose your data, you can*

go to the data centre and retrieve your data. Through the research support centre, we encourage researchers to go through the research support centre but I cannot blame researchers for not using the facility because we have limited capacity at the research support centre; we only have one data officer and we have got so many research projects in the college and the data officer cannot cope.

For UNI1, it was noted that there was no notable infrastructure dedicated to research data management. The Director of Research at UNI1, however, explained that plans were underway to put in place such infrastructure. It was revealed that the university was in the process of setting up a research ethics committee and as part of this process, the university intended to come up with mechanisms for managing data including establishing a proper data infrastructure. For example, it was revealed during the interview that all researchers obtaining ethics clearance from the university's ethics committee would be required to submit their data for deposit into the university's data repository after completion of their research projects. The only available storage facilities used by researchers, according to the Director of Research included personal laptops, office computers, external hard drives and flash disks. According to the Director of Research, plans were at an advanced stage for the university to build a new library and the library design included a component of a data centre. The Director of Research at UNI1 commented that:

At the moment the answer is no, we don't have but what I can say is that we have plans and I would like to link to efforts that are underway. I think we are at advanced stage of establishing an ethics committee. Now, as part of that system, we will not only be collecting proposals, we are going to create a database for proposals and when people go out to do their research, they will also be submitting reports including data. So, we are planning to put in place mechanisms where we can keep our data and this will be the entry point for data sharing; we will put it as a requirement that deposit your data as well and make those procedures for re-use. And as we build the database, I think the infrastructure will be available but at the moment, we have limited infrastructure, researchers are only using office computers and their personal laptops. In the meantime, we will collaborate with the library, we will come up with a system [which] apart from depositing research publications, data should also be deposited in the institutional repository. After carrying out research approved by the university ethics committee, researchers will be required to deposit their data in this particular database.

These results align well with those realised from library staff and researchers. As can be noted in section 5.2.3.1, the majority of library staff indicated that the common digital storage facilities at the university that researchers could use to store data included flash disks, computers, email accounts and external hard drives. It was further noted in section 5.2.5 that most library staff at both universities indicated that one of the challenges affecting research data management was insufficient storage and network infrastructure.

On the part of researchers, it was reported in section 5.3.2.5 that a greater number of researchers used external hard drives and emails to share their research data. These findings are further confirmed by those reported in section 5.3.3.4 where it was revealed that the common digital storage facilities used by researchers at both universities included personal computers, office computers external hard drives, freely available software or services such as Google Drive, flash disks and email accounts. Claims by the Director of Research at UNI2 that researchers lost data when digital storage devices crash support findings reported in section 5.3.4.1 where it was noted that most researchers indicated that they lost their data due to accidental damage of storage facilities. Claims by the Director of Research at UNI1 that the university had not put in place any infrastructure for managing data support findings reported in section 5.3.4.2 where the majority of UNI1 researchers indicated that some of the challenges that affected their research data management activities included lack of storage infrastructure and lack of support from the university.

5.4.4. Research data management skills

One of the themes addressed during the interview with directors of research was adequacy of skills for researchers in managing research data. The researcher wanted to know if these offices supported researchers to acquire the right mix of skills for managing their research data. The Director of Research at UNI2 indicated that the personnel working in the data centre had adequate skills for managing data generated by some researchers at the college. However, the Director of Research was not sure if researchers within the college had enough skills to manage their data because the college had not organised any training sessions or workshops on the same. The Director of Research at UNI2 commented that "As I said, we have people who are dedicated for storage, analysis, we have people in research support centre who can do the analysis. My office here coordinates and makes people aware that we have these services". For UNI1, the Director of Research indicated that the university had not conducted any workshop in research data management implying that researchers were not exposed to these skills. The

Director of Research indicated that once the university established a database or a repository for preserving data, it could work with the Office of the Director ICT at UNI1 to conduct some training to orient researchers on how to use the facility. In this respect, the Director of Research commented that:

> *…when it is implemented, we will work with the ICT Directorate in training users. Of course, my office will initiate the awareness campaigns and we will fund to make these things known by researchers. In short, the Directorate [Research Directorate] will fund these activities meaning it's my office that is responsible. So, yes, we will fund for training for researchers.*

Results in this section help explain the reason why quantitative data from both library staff and researchers showed that they lacked skills in research data management. Specifically, it was noted in 5.2.3.2 that the majority of library staff lacked skills in helping researchers decide which data is important to preserve; deciding which data can be safely shared; determining standards for identifying sensitive data; helping comply with licenses, regulations and mandates on data management; assigning metadata; and determining data storage and preservation on long term. The problem on the part of librarians was further noted in section 5.2.4.2 where the library staff at both institutions need training in identifying new standards and practices for curation; curating digital objects using curation lifecycle; long term digital data preservation strategies; selecting digital objects for preservation; storing digital information using standards; citing and transforming data. The problem of skills in research data management was further noted in section 5.2.5 where librarians indicated that they lacked curation skills and training in research data management.

These results also give support to those realised from researchers as reported in section 5.3. To begin with, the findings in section 5.3 showed that most researchers at both universities did not attend any workshops or training in relation to research data management. It was revealed in section 5.3.2.6 that researchers were discouraged from sharing data because they lacked data sharing skills. In section 5.3.3.7, it was noted that 141 (74.2%) researchers had not attended any workshop in research data management. It was further noted in section 5.3.3.9 that most researchers were not competent in various research data management activities including writing data management plans; advanced computing; short term data preservation strategies; long term data preservation strategies; preservation planning; identifying standards and practices creating preservation metadata; depositing data into repositories or archives; adhering

to data management standards; disposition of data; and migrating data to newer file formats. Finally, it was revealed in section 5.3.4.2 that some of the challenges that researchers faced in managing their data included lack of curation skills and training; lack of guidance and support; and lack of skills in sharing data and lack of skills to create metadata.

5.4.5. Factors affecting research data management

The last item in the interview guide focused on finding out from directors of research the factors that affected the management of research data in the two universities. According to the Director of Research at UNI2, the key challenge was lack of investments by the college in data curation. The college had not invested in infrastructure such as servers and power back-ups. The results revealed that although the college had established a data centre, generally, the college had shortfalls in critical research data management areas and worse still, the staff in the data centre were not enough as there was only one data officer offering research services to hundreds of researchers at the college. Another challenge according to the Director of Research at UNI2 was lack of publicity about the services offered at the data centre. An extract from an interview with the Director of Research at UNI2 about factors that affected research data management at the college is as follows:

> *Number one, I can say lack of investment, you need good servers and proper back up of power. Number two, is also personnel, skilled people who can really, you know, convince researches about good data management practices. And as well as, maybe we don't really publicise the activities that we are doing in the research support centre and its data centre. The awareness part is not really done to a scale that people can be really aware of research services.*

A number of factors that affected research data management at UNI1 were highlighted by its Director of Research. The first challenge was that the concept of research data management was a new one and many researchers had not bought the idea of data sharing and re-use. The other key challenge was infrastructure. It was revealed that the university did not have any storage facilities and instead researchers were only using office computers, their personal laptops and emails. It was further revealed that the university did not have personnel who could manage data generated by researchers. According to the Director or Research, the only viable short term option was to make use of the library staff but the challenge was convincing the library management to assign a member of staff to assume the position of a data officer. Again, a concern of the Director of Research was that even if the library was willing to sacrifice one

of its staff to the position of data officer, there was still a need to invest in training because being a new concept, data curation needed new skills and competencies. Another challenge highlighted by the Director of Research was coming up with proper data management policies that could guide preservation, sharing and re-use of data; and policies that could stipulate many aspects such as how to reward those researchers who shared their data for re-use by others, data access (by internal and external users), data appraisal and data disposition. Below are a selection of some extracts from the Director of Research at UNI1.

> *I think the first that I immediately think of is the one that I have already highlighted, its new issue, so buying this idea by the people who are generating data and accept the concept of sharing will require a little of some effort. Convincing people that you can use data generated by others and again also convincing generators of the data that they need to share their data just as they do with their publications. It may be a challenge because again it's a new concept, will the library be willing to be engaged on fulltime basis to be managing data; they are doing it for the repository [institutional repository] you know for our digital works, maybe the same person will be assigned to manage data? But now, data is a different thing as it may require different technical expertise. I think beyond that, I know everybody will expect something from the use of the shared data. Even for the university; for internal use, we can have some policies but assuming somebody from outside wants to use our data, there we need to be careful because it will be a challenge. What if the university wants external users to pay for our data but then some may not want to pay because this is a publicly funded institution?*

The results are supported by those reported in section 5.3.3.6 where it was noted that researchers were of the opinion that for research management activities to run smoothly, there was a need for universities to offer infrastructural support in various ways such as establishing a process for managing data for a period of five years or more, establishing technical support for managing data and establishing funds to support data management. Lack of policy frameworks was also mentioned by researchers in sections 5.3.2.9 and 5.3.4.2 as one of the factors affecting data management amongst researchers in the two universities. As further reported in section 5.3.4.2, lack of policies at both institutions could have influenced researchers at both institutions to state that they found it difficult to access and re-use data produced by fellow researchers within their universities. In addition to the preceding reason, failure by UNI2 to publicise services offered at the data centre may have contributed to its researchers mentioning that they found it difficult to find data for re-use purposes.

Failure by the universities to provide enough infrastructure and training as mentioned by directors of research also explain the reason why the majority of researchers indicated in section 5.3.4.2 that one of the challenges they faced in their research data management activities was lack of support from the university. Again, failure by universities to provide research data infrastructure may have been a contributing factor for all 103 (100) researchers at UNI1 and 84 (100) researchers at UNI2 to resort to the use of personal computers as revealed in section 5.3.3.4 and also for all these researchers (100%) to state that they never used the institution's available networked capacity. It can then be concluded that the 22 (61.1%) librarians who said that they already helped researchers to store copies in a data repository or archives (as reported in in section 5.2.6.2) may have meant helping researchers store copies of their research publications in institutional repositories which existed at both universities.

5.5. Summary of chapter

This chapter has presented two forms of data about research data management at two public universities in Malawi. The first part presented quantitative data collected from library staff and the second part presented another set of quantitative data gathered from researchers. The final part of the chapter analysed data from two directors of research, one from each of the universities. Data was presented based on various key themes that guided the study namely; data creation, sharing and re-use practices; research data preservation practices; research data management competencies; and challenges that affected the management of research data. This section provides a highlight of some outstanding findings coming from the study and sets the tone for the subsequent two final chapters.

Both universities had put in place mechanisms to motivate researchers to conduct research and publish the findings in creditable journals. Some motivating factors included promotion to senior ranks, monetary rewards and sponsorship for the presentation of findings at conferences. The key data storage and sharing facilities were mainly personal laptops, emails, and flash disks and external hard drives.

Generally, results revealed that the concept of research data management was new to both universities. As a result, various elements of the concept were not embraced by the universities. However, universities appreciated benefits associated with research data management such as advancements in science, affordable costs associated with data re-use and convenience on the part of data re-users. Findings showed that the majority of researchers were not sharing their data nor re-using other researchers' research data mainly because of lack of policy frameworks

and skills in research data management. Researchers were however, interested in sharing their data on condition that they put restrictions over access of the data they shared.

The key challenges in research data management resulted from the fact that universities had not invested in research data management. For example, researchers lacked skills and competencies in basic various aspects of research data management because no training workshops on research data management were organised by universities. Apart from personal laptops, emails and external hard drives, universities had not put in place centralised systems for managing research, save for UNI2 which had a data centre which nonetheless lacked capacity and operated on a small scale.

Findings of t-test and ANOVA showed that there were statistically significant differences between the two universities in terms of data sharing tools, factors that affected data sharing, conditions for sharing data, data re-use factors, data storage facilities, data back-up strategies, support required from universities and loss of research data loss. On the other hand, t-test and ANOVA showed no statistically significant differences between the two universities in competencies or skills for managing research data and factors that affected researchers' data management.

In the process of presenting qualitative data, the researcher triangulated it with quantitative data from library staff and researchers in order to lay a foundation for discussion of findings in next chapter, that is, Chapter Six.

CHAPTER SIX

DISCUSSION AND INTERPRETATION OF FINDINGS

6.1. Introduction

This chapter aimed at discussing and interpreting the findings obtained from the quantitative and qualitative data presented in the preceding chapter (Chapter Five). The researcher draws wisdom from Creswell (2013) and Lunenburg and Irby (2008) who advise that discussing and interpreting the findings involve providing meaning to those results by linking them with the research objectives, theoretical frameworks and the existing literature. Particularly, Lunenburg and Irby (2008) warn researchers not to deviate the discussion and interpretation from the data. "Stay close to the data" but use sparingly the technical details of the analysis such as *F* and *p*-values because they mostly serve their role in the results chapter, say Lunenburg and Irby (2008, p. 229). Similarly, the University of the South California (2019) maintain that it is worthless and time wastage restating the results in the discussion chapter; instead, 'bridge sentences' should be used to relate the discussion to the results. In other words, reference to specific data should only be made to support particular statements in the discussion (Lunenburg & Irby, 2008). This means that to write an effective discussion, the researcher needs to be thoroughly knowledgeable of the study's data and results, theoretical frameworks and the extant literature; these are the cornerstones of the discussion. In this chapter, the researcher attaches meaning to the results reported in Chapter Five by interpreting them through the lens of the theoretical frameworks which underpinned the study and by contextualising the findings into the fold of the existing related literature on research data management which was discussed in Chapter Two. The adopted models included the CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008). These models are comprehensively discussed in Chapter Three.

The aim of the study was to investigate the research data management practices at two public universities in Malawi. Four research objectives were formulated in Chapter One (see section 1.5.1) to guide the study namely, to determine research data creation, sharing and re-use practices in public universities in Malawi; to investigate research data preservation practices in public universities in Malawi; to investigate competencies that librarians and researchers needed to effectively manage research data in public universities in Malawi; and to find out the challenges that affected the management of research data in public universities in Malawi. This chapter is organised in line with the sequence of research objectives mentioned. Apart from

demographic characteristic of respondents, the discussion is anchored in these research objectives.

6.2. Demographic characteristics of the respondents

In the CCMF, Lyon et al. (2011) place an emphasis on understanding characteristics of RDM stakeholders that are essential in examining the readiness of the communities to perform RDM activities. In the light of this, the researcher examined demographic details of the participants in terms of gender, qualification, rank, and university affiliation.

In terms of university affiliation, there were more respondents from UNI1 in both categories of quantitative data. This is perhaps because while UNI1 is a fully fledged standalone university, UNI2 is a college under the University of Malawi. It is therefore, not surprising that while UNI1 had six faculties with diverse academic specialities, UNI2 had only three with specialities in health sciences. Across the participants, there were more males than females. A possible explanation is that higher education in Malawi is dominated by males, a reason offered by Chawinga and Zozie (2016) and Chipeta, Dube, Chawinga, Malemia, and Chaura (2018) who observe that historically, universities in Malawi enrol more males than females mainly because of cultural connotations which force females into early marriages thereby abandoning school at a young age. On the contrary, males are considered as the main source of income who should be supported to excel in school so that they can render support to their family members (Chipeta et al., 2018). It is therefore expected that poor enrolment levels of female students would translate in to fewer females being recruited for any position in the university.

In Malawi, teachers were offered good pay to entice them to leave the teaching profession where they worked as secondary school teachers to join the LIS profession as librarians in university libraries; considering that these teachers had no any formal education in LIS, they were then sent to pursue LIS studies (postgraduate and masters' degrees) in the UK and Botswana. Both men and women were equally enticed to join the profession but as already alluded to, the fact that few female students enrol in universities in Malawi, it is unsurprising that fewer females are recruited in the LIS profession. In contrast, the USA librarianship is a female majority profession; as of 2010, 82.8% of all working librarians were women (American Library Association (ALA), 2011, p. 2).

The qualifications of library staff manifested a pyramidal shape: at the top of the pyramid, there is a small number of staff with a PhD and a Masters; in the middle, there were those with a sizeable number of Bachelor's Degrees and the bottom of the pyramid had the highest number

of staff with a Diploma and a Library Certificate. Independent t-test results also confirmed that there were no statistically significant differences in qualifications between the two universities. On the part of researchers, most of them had a Master's Degree, followed by those with a PhD and only a few had a post-PhD. However, the study noted that UNI2 had more staff with a PhD than those with a Master's Degree while UNI1 had more researchers with a Master's Degree than those with a PhD. The results were further confirmed by performing an independent t-test which revealed that there were statistically significant differences in terms of qualifications in favour of UNI2. This means that UNI2 had a fair distribution of qualifications across the variables while UNI1 had a skewed distribution of variables dominated by Master's Degree qualifications. It is understandable that many researchers had a Master's Degree as their highest qualification because according to the National Council for Higher Education (NCHE) (2018), a Master's Degree is the minimum qualification to teach at an institution of higher learning in Malawi. However, with the introduction of postgraduate studies at UNI1 and UNI2, lecturers are being challenged to acquire a PhD and post-PhD.

There were more library staff at the rank of senior library assistants followed by those at the ranks of library assistants, senior assistant librarian and college librarian. No statistically significant difference was noticed in terms of rank at the two universities as revealed by the independent t-test. Across the two universities, there were more staff at lecturer level than other ranks such as senior lecturer, associate professor and professor. However, UNI2 had more professors and associate professors than UNI1 and the independent t-test results also confirmed that there was a statistically significant difference in favour of UNI2.

6.3. Research data creation, sharing and re-use practices

The study sought to identify research data creation, sharing and re-use practices in Malawian public universities focusing on research output, data format(s), data sharing practices and data re-use practices.

6.3.1. Research output and its importance

To better understand research data management activities, it is fundamental to understand where data comes from (Higgins, 2011; Research Information Network, 2008). Thus, to underscore the importance of data in the research data life cycle, both the CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008) integrate data creation activities. The two universities were already involved in intensive research which they valued highly because of three key reasons. Findings from interviews with directors of research showed that

research helped to drive the national economy, contributed to the international ranking of the universities, and it was the source of funds for the universities. The majority of researchers had their papers already published in various journals and other conventional scholarly outlets. Comparatively, all 84 (100%) UNI2 staff had published papers against the 94 (91.3%) UNI1 counterparts. According to the CCMF (Lyon et al., 2011), open communication of research methods and results contribute to scientific progress. It can therefore be concluded that, considering that most researchers in the present study had published their research results in creditable journals (mostly identified with the help of librarians), the universities were contributing to the advancement of science at institutional, national and international levels.

The factor of university ranking as one of the driving forces for conducting research at the two universities cannot be disregarded. According to Aguillo, Bar-Ilan, Levene and Ortega (2010) and Soh (2015), one of the key parameters that different ranking systems such as the popular QS World University Ranking and Times Higher Education World University Rankings take into account is publication and citation counts. In that regard, the two universities participated in research publication activities to remain visible and relevant within the competitive international education framework; that is the only way they can start or continue attracting exceptional students, distinguished academics and the much cherished donor research grants. Findings of the present study show that research propelled the national economy; this notion is supported by Woolfrey (2009) who observes that both text publications and secondary data provide policy makers with evidence-based information for formulating policies which propel national, regional, and international economies.

6.3.2. Data formats

With regard to the element of *conceptualise* of the DCC Curation Lifecycle Model, Higgins (2008) highlights the importance of proper planning in how the data will be curated including captured. In view of this, when investigating RDM, it is important to understand the formats in which it is created and captured (Higgins, 2011; Research Information Network, 2008; Scott, 2014, p.121; Walters & Skinner, 2011) because data can vary extensively from one research discipline to the other (Krier & Strasser, 2014; Ohaji, 2016, p. 25). The most common forms of data generated by researchers at both universities in the present study included digital texts, audio recordings and spreadsheets. The less common types of data include spatial data, computer codes and video recordings. The only type of data which was significantly common at both universities was audio recordings. The fact that 81(78.6%) UNI1 and 81(96.4%) UNI2

researchers produced data in digital text gives credence to observations by Cox and Pinfield (2016), Kahn et al. (2014), and Ohaji (2016) that the evolution in ICTs in universities and research institutions has contributed to the explosion of research data generated in digital format. Audio recordings are very popular across research fields dealing with human beings while specimens are particularly common in medical sciences and these results present no surprises. UNI1 being a university dominated with social sciences programmes conducts qualitative research whose data is mostly collected using recorded interviews. In addition to collecting samples commonly called specimens - which were common at UNI2 (81 or 96.4%), researchers in medical sciences such as UNI2 commonly conduct interviews with participants as complementary to samples examined in laboratories.

The researchers at neither university produced data in the form of video recordings and artistic products. Using video recordings in human research presents an ethical dilemma; many participants may not be willing to be filmed when providing responses to researchers. Hence, data in form of video recordings in human research is somewhat rare. The CCMF (Lyon et al., 2011) warns that some ethical obligations might limit what researchers can do with the data apart from the core purpose for which consent was sought and granted by the participants. Likewise, Chen and Wu (2017) also noted that none of the chemistry researchers in China used video recordings. However, unlike the present study, Chen and Wu (2017) found that the researchers did not use audio recordings. The difference can be attributed to chemistry researchers focussing more on experiments that rarely involve direct interaction with human beings.

6.3.3. Data sharing practices

The DCC Curation Lifecycle Model (Higgins, 2008) articulates in its *access, use and reuse* action that users and re-users should have access to data and use particular data on a daily basis depending on the restriction access and use conditions imposed by creators. Data can only be accessed, used and reused if creators are willing to share their data sets. In this regard, the study investigated data sharing practices focusing on data sharing motivating factors, data sharing tools and factors affecting data sharing.

Factors that motivate researchers to share research data

Guedon (2015) and Wicherts and Bakker (2012) reason that for ages researchers have been conditioned to sharing their papers or research findings but are now challenged to share their data too. The study revealed that the majority of researchers (113 or 59.5%) shared or were

willing to share their data. The essential factors that encouraged researchers to share data at both universities included personal initiatives and journal policies. Research funders, university policies and open access initiatives did not compel researchers to share their research data. The DCC Curation Lifecycle Model (Higgins, 2008) underscores that institutional RDM policies are of critical importance because they provide a framework for undertaking curation activities such as capture, appraisal, description, preservation, share, access and reuse. The reluctance of researchers to share data due to the absence of institutional policies in the present study warrants UNI1 and UNI2 research stakeholders seriously consider this aspect.

The final home of most research findings is the journal article. The study found that a good number of researchers were compelled by journal policies to share their data. Most journal publishers of high repute, namely *Atmospheric Chemistry and Physics*, *F1000Research*, *Nature*, *ScienceDirect* and *PLOS One* have formulated and implemented policies that require authors to submit manuscripts alongside the data sets (Bond-Lamberty, 2018; Fecher et al., 2015; Savage & Vickers, 2009). The only stumbling block, however, is that these journal policies have been criticised for being weak and not properly enforcing the requirement and this could be the reason that not all researchers in the present study attributed their data sharing practices to this factor. A study by Wiley (2018) noted that 76% of engineering journals had weak research data-sharing policies implying researchers could choose to violate the policy requirements with impunity.

Open access movements advocate for open data access so that other scientists and health experts can access the latest evidence, draw on it to advance their own research, and benefit from this knowledge (Mundel, 2014). Guedon (2015) is of the opinion that for many centuries, researchers have learned to share their papers or research findings; now they must learn to share their data. However, open access initiatives did not influence researchers to share data; 174 (93%) researchers were not compelled by this factor to share data. The result could be attributed to the absence of workshops through which issues of depositing data in open access repositories could have been raised, discussed and initiatives taken to popularise them. The development is unfortunate because, according to the Berlin Declaration (2003) and the European Commission (2012), depositing research data in open access data repositories is the only straightforward and sure way of ensuring access to research data. More importantly, the DCC Curation Lifecycle Model (Higgins, 2008) highlights that research data should be made available to designated users or re-users; the findings are therefore, in contradiction to the requirements put forward by this model.

While these findings support those of prior studies, they also contradict them in some ways. For instance, although research funders are hailed for compelling researchers to share data (Charbonneau, 2013; Chen & Wu, 2017, p. 346), this factor had little impact on the current study. Based on the findings, an unrelated reason can be attached to each of universities under study. According to the findings, the majority of researchers at UNI1 conducted self-sponsored research hence they were not compelled by research grant bodies to share their data. On the other hand, considering that research at UNI2 was mostly funded by external research funders, the non-compliance of funder's demands could be attributed to weak data sharing policies and failure by research grant organisations to enforce these polices. The results are not unique to UNI2 because in the USA, despite the National Research Foundation of the USA adopting a research data sharing policy that requires its research grants recipients to share their data (Cohn, 2012), the policy is hardly stringently enforced (Borgman, 2012). In addition, by virtue of using the college name and facilities to conduct their research activities, all data produced by UNI2 researchers (whether self-funded or externally funded) belongs to the university. However, this was a mere directive and not a policy because the college is yet to adopt an RDM policy.

Although the Malawian Government established the National Commission for Science and Technology to oversee research activities including funding research and supporting formulation and enforcement of research polices in universities and researcher institutions, the institution does not offer research funds to researchers as is the case with other similar national research organisations such as the National Research Foundation of South Africa (Chiware & Mathe, 2016, p. 2; Koopman, 2015; Matlatse, 2016) and the National Science Foundation of USA (Cohn, 2012; Borgman, 2012). Instead, activities of the institution have been confined to ethics clearance issues because it operates on an inadequate and unsustainable budget. Hence, enforcement of data sharing at national level is unsurprisingly problematic. In their study in Uganda and Tanzania about data sharing practices in health sciences, Anane-Sarpong et al. (2017) observe that RDM activities are slow and unsatisfactory in Africa due to financial constraints. Results of the present study are unsurprising because, according to the CCMF (Lyon et al., 2011), data-intensive research requires some degree of investment in major studies such as funding research projects which is presently lacking at the two universities.

Data sharing tools

The passion by researchers to share data with their peers or other users can only become a reality if they have access to proper data sharing infrastructure. These data sharing tools are part of the complex technical infrastructure proposed by Lyon et al. (2011) and Higgins (2008) in their CCMF and DCC Curation Lifecycle models respectively. According to the CCMF and the DCC Curation Lifecycle models, technical infrastructure is necessary for preservation, data discovery, access and collaboration. Some common platforms that researchers can use to share data include "attaching data sets to published articles, depositing data sets in repositories, posting data on a personal or laboratory website, or fulfilling requests from other researchers for data" (Wallis et al., 2013, p. 2). In the present study, findings show that the common research data sharing tools were external hard drives and emails. The reason researchers used these two tools is that these were provided by their universities. It has to be acknowledged that, as will be noted in section 6.4.2, these are the same tools that researchers mostly used to store their data. Emails seem to be popular data sharing tools because Rowhani-Farid and Barnett (2016) also noted that most researchers used emails to share data. The present study established that researchers at both institutions did not use social networks, blogs/wikis, journal websites, funder's websites, university websites, clouds, the principal investigator's website, the national network, or regional network.

These findings are contrary to results of some prior studies. Shakeri (2013) reports that at Kent State University in the USA, the majority of researchers used cloud-based password applications namely, Dropbox, Evernote, and Google Drive to share data with their peers. Failure by researchers in the present study to use web-based applications such as Google Drive can be attributed to the lack of knowledge and skills especially considering that both universities offer email facilities powered by Google of which Google Drive is one the services offered. Researchers could not use the national network as it does not exist in Malawi. Absence of such an infrastructure is in contrast to the requirements proposed by the CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008) which encourage research stakeholders to build a robust infrastructure for capturing, appraisal, description, preservation, access, sharing and security of data. Considering that UNI2 has established a data centre, the expectation was that its researchers could store and share data using this platform. Instead, most researchers did not use this facility because, in addition to its limited capacity, its services were not well publicised among researchers.

Factors that discourage researchers from sharing research data

Generally, factors affecting data sharing were not influenced by universality affiliation as demonstrated by ANOVA which revealed a weak similarity of the factors between the two universities.

As noted in section 6.3.3.1, not all researchers shared their research data mainly because most of them lacked data sharing skills. The reason is that the concept of research data management is new in Malawian universities according to the interviews with directors of research. A quote from the Director of Research at UNI2 reads: "it's a new concept of course, but I understand it as where researchers from different institutions can share the data as well as you can re-use that data for future discoveries". Being a new concept, the universities had not organised any training workshops to equip researchers with RDM skills; the study showed that 141 (74.2%) researchers had not attended any such workshops. It appears that the concept of data sharing is not only new to the Malawian research environment because Matlatse (2016) also describes it as an emerging concept in South Africa. Similar findings were reported by Tenopir et al. (2011) who noted that 59% researchers in the USA indicated that their universities did not equip them with best practices for data management hence, they lacked RDM skills.

Another challenge was metadata and their standards. In the DCC Curation Lifecycle Model, Higgins (2008) explains that the activity of description and representation of information is fundamental for long-term access to data. The activity involves assigning various types of metadata (administrative, descriptive, technical, structural and preservation) to data using appropriate standards (Higgins, 2008, p. 134). According to the findings, researchers did not share their data because it was not documented meaning the data lacked proper description (metadata). Hence, the data was not widely readable. It then makes sense that the absence of metadata influenced readers to point to a lack of standards or guidelines for managing data. Failure to document or assign metadata can also be attributed to the lack of skills by researchers at both universities. Well documented data is essential for data sharing and re-use (Enke et al., 2012; Nelson, 2009; Tenopir et al., 2011; Yoon, 2015, p. 144; Yoon & Schultz, 2017). This observation is echoed by the CCMF (Lyon et al., 2011) which states that *de facto* standards (data formats, data collection methods, processing workflows, data packaging and transfer protocols, data description, semantics, ontologies and vocabularies, and data identifiers) developed by research communities need to be shared and understood by researchers within a particular research field. In this context, researchers did not share their data because they could

not assign such metadata and standards implying that even if they shared such data, re-users could not comprehend and effectively use it.

The study found further that 160 (85.6%) researchers did not share data because they lacked time. Researchers are busy people with a workload traversing teaching, supervising students, conducting research and partaking in community service. It is hardly surprising that they have limited time to share their data with peers and the public. Similar findings have been reported in the extant literature. In the USA, Tenopir et al. (2011) found that 53.6% of researchers were unable to share data due to insufficient time. An argument can be put forward that data sharing should be considered a natural extension of research activities, hence the issue of rewards should not be entered into the equation. However, unlike sharing published texts through journals or other outlets, publishing data comes with no innate rewards (Acord & Harley, 2012). This line of thought is well supported by the findings of the current study where 96 (95.2%) researchers at UNI1 indicated that they did not share data because there were no rewards to incentivise them. Similarly, Tenopir et al. (2011) also found that 59% of researchers confirmed that they did not share their data because there were no monetary incentives; they were not provided with funds to support data sharing undertakings. Both models informing this study did not clearly address the implications of a lack of time on RDM. Perhaps, revisions of these two models or framers of related new models need to consider this factor. The present study was inspired to propose an RDM model incorporating aspects missing in the previously developed models.

Conditions for sharing research data

Encouraged by Tenopir et al.'s (2011) study in the USA, the current research investigated conditions that motivated or had the potential to motivate researchers to share data. Placing restrictions on the data they shared (174 or 93%) and having their data cited by re-users (176 or 94.1%) were the key conditions that would inspire researchers to share data at both universities. Researchers' insistence on putting restrictions on the data they share can be attributed to one principal factor. With increasing cases of data misuse (Doorn et al., 2013; Fecher et al., 2015, p. 16), researchers in the study were mindful of possible misuse or misappropriation of their data by re-users. Researchers do not want their data to be used for commercial purposes, for example. The common forms of data misuse include falsification, commercial misuse, competitive misuse, flawed interpretation, and unclear intent (Doorn et al., 2013; Fecher et al., 2015, p. 16). Fears of data misuse as a deterrent to data sharing have also

affected researchers in South Africa, reveals a study by Denny et al. (2015) who explored data sharing practices in that country. It was noted that researchers were hesitant to share their data because they feared potential users might misuse it through commercial exploitation; they were further afraid that re-users might misunderstand their data thereby arriving at incorrect conclusions that could endanger their research integrity (Denny et al., 2015). As one strategy of diffusing fears of researchers in Kenya about data misuse, Jao et al. (2015) reason that it is necessary to build trust between primary data generators and potential data re-users through awareness campaigns on the ethical use of secondary data.

These findings are similar to those of Tenopir et al.'s (2011) study in the USA which also revealed that 64.4% of researchers were willing to share their data if they placed access restrictions on it. Tenopir, Sandusky, Allard, and Birch (2014) appreciate that the willingness to share or not to share data is mostly personal and some key factors that influence them include privacy concerns, publishing opportunities, and the desire to retain exclusive rights to data. Similarly, in Germany, Fecher et al. (2015, p. 16) found that due to fears of misuse and misappropriation scientists insisted to know the users of their data before accepting sharing. On the need to have their data cited, researchers want secondary users of their data to cite it as part of acknowledgement, since data sharing comes with no apparent rewards, citation becomes the notable way researchers can be recognised for their data sharing efforts. Along the same lines, researchers at both universities were not willing to place their data in repositories without restrictions. These results reflect the DCC Curation Lifecycle Model (Higgins, 2008) which warns re-users beforehand to be mindful of access and re-use conditions imposed by the primary creators of the data.

6.3.4. Research data re-use

The DCC Curation Lifecycle Model (Higgins, 2008) places emphasis on access, use and re-use of data by both designated users and re-users, hence the need to ensure that data is readily accessible for this purpose. The findings show that 108 (57.8%) researchers used or were willing to use secondary research data generated by other researchers or research institutions. The findings from interviews with the directors of research revealed some benefits associated with data re-use. They included advancements in science, affordable costs associated with data re-use and convenience on the part of data re-users in regards to issues of ethical clearance.

Prior literature informs the scholarly community about the many benefits of data re-use. Shakeri (2013) observes that re-use of data can minimise the cost and redundancy of research

data production and this observation is confirmed in the present study as commented by the Director of Research at UNI1 that " people must appreciate that it's cheaper to use data already generated than start afresh collecting the same data" and in agreement, the Director of Research at UNI2 said "You know conducting fresh research is costly but students at that level, at masters and undergraduate, they can just go into a database and re-use that data and that becomes much easier - it's less costly and there are less hassles in ethical clearance". According to Tenopir et al. (2011, p. 1), "when data is available, (re-)collection of data is minimised; thus, use of resources is optimised". It is for this reason that Piwowar (2011) concluded that data reuse is sensible financially. This finding resonates with the finding of a study conducted by Ng'eno (2018) who equally found that 88.7% of researchers in some Kenyan research institutes benefited from data sharing because it encouraged scientific enquiry and debates, and also reduced the cost of duplicating data collection.

The issue of ethics clearance emerged in this study and is worth discussing. The findings showed that researchers at UNI2 encouraged the re-use of existing data thereby causing few or no worries or problems on the part of re-users especially undergraduate and post graduate students. This notion has been propagated before by Woolfrey (2009) who notes that some researchers, especially students, tend to struggle to obtain permission from authorities or organisations to conduct studies. Hence, according to Whitlock (2011), students can be directed to previously published data sets on a topic of interest and have the opportunity to extract and use the data in their research activities.

The advancement of science was another factor influencing data sharing at both universities. The re-use of data generated by other researchers and research institutions could enrich science. These findings support the notion that previously generated data is the cornerstone of science (Wallis et al., 2013). This is achieved through the re-analysis of data generated by other researchers thereby arriving at new breakthroughs (Takashima et al., 2018; Tenopir et al., 2011). In addition to arriving at new knowledge, data sharing enhances science by policing research fraud. Through re-analysis, data is verified to examine its accuracy and reproducibility thereby safeguarding against possible misconduct related to data fabrication and falsification (Doorn et al., 2013; Elsayed & Saleh, 2018; Fecher et al., 2015; Watson, 2015) which are said to be on the increase in the research community (Doorn et al., 2013). In support of the findings, a similar study in Germany sums up the role of data-reuse in regard to propelling science; Fecher et al. (2015, p. 14) report that data re-use makes "research better", it is a basis for

"feedback and exchange", provides "consistency in measures across studies to test the robustness of effect" and enhances "reproducibility of one's own research".

Factors that affect data re-use

Although the study has proven that data re-use comes with various benefits for the field of science, not many researchers re-use data due to a number of factors. According to the findings, the factors that discouraged researchers from using data generated by others included difficulty in accessing reusable data, lack of metadata (metadata standards) and legal/ethical restrictions. On the one hand, some factors, namely difficulty in integrating data, not trusting other researchers' collection methods, data may be misinterpreted due to its complexity, and data may be used in other ways than intended, did not discourage researchers from re-using data. The independent t-test revealed a strong similarity of responses between the two universities in the factor of legal/ethical restrictions implying that the factor equally affected both universities.

How the lack of access to re-usable data frustrates researchers from re-using data is well documented in the literature. All 187 (100%) researchers strongly or somewhat agreed that they found it difficult to access previously produced data for re-use purposes. Yet, according to the DCC Curation Lifecycle Model (Higgins, 2008), data should be accessible to both designated users and re-users. A few previous studies have reported similar results. In a study conducted in the USA involving 1,329 researchers, 60% of the researchers acknowledged that they could not use data produced by other researchers because the data was unavailable or inaccessible (Tenopir et al., 2011).While lack of data sharing infrastructure and incentives are the contributing factors for the unavailability of re-usable data in the current study, Tenopir et al. (2011) report that in in the USA which is a developed country, the key contributing factor is absence of rewards to encourage researches to make their data publicly accessible. The results suggest that establishing a sound reward system and a robust data sharing infrastructure should be considered by UNI1 and UNI2.

Metadata are critical for data discoverability and re-use (Brown, et al., 2015; Tenopir et al., 2011; Woolfrey, 2009; Yoon, 2015; Yoon & Schultz, 2017; Zvyagintseva, 2015). According to the CCMF (Lyon et al., 2012), for data to be accessible by designated and potential re-users, various data features such as formats, collection methods, description, and data packaging and transfer protocols should be well documented. Therefore, interpreting the findings of the present study using the CCMF (Lyon et al., 2012), it can be concluded that researchers found

it difficult to use previously generated data which did not contain metadata as they could not understand the context in which the data was created, the purpose of the data, creators of the data and legal aspects associated with the data. The present study found that where metadata was assigned to data, the metadata was not standardised. Brown, et al. (2015), Musgrave (2003, p. 8), Woolfrey (2009) and Yoon (2015) observe that the lack of standardised metadata creates interoperability problems. Absence of standardised metadata suggests that when primary investigators share their data, secondary users find it difficult to understand and re-use it. In that regard, the current study therefore supports previous observations in the literature that detailed and standardised metadata can stimulate researchers to develop more interest in data re-use in their research activities. Standardised metadata facilitates easier exchange and use of data sets across multiple organisations, systems and platforms, argue Abbott (2009) and Shakeri (2013).

On a positive note, the current study shows that researchers had trust in data produced by other researchers and would not misinterpret it as long as it is accompanied by comprehensive metadata. Trust is important in data re-use. A study by Yoon (2015) involving researchers in the USA revealed that researchers were more inspired to use data generated by other researchers if they were convinced it originated from trusted primary generators. To establish the originality and authenticity of data, metadata becomes inevitable because they are the key pointers to primary producers of data. In the current study, researchers indicated that they could not misinterpret data produced by other researchers suggesting that they have adequate data analysis skills and competencies. In this regard, comprehensive and accurate metadata is vital because it informs re-users of data of more features of data in terms of participants and variables measured. It is for this reason that the DCC Curation Lifecycle Model (Higgins, 2008) informs curators to undertake preservation actions that ensure authenticity, reliability, usability and integrity of data are maintained by assigning comprehensive metadata (administrative, descriptive, technical, structural and preservation). Hence, it becomes easier for the users to create parameters for re-analysis and interpretation using any data analysis tools. In summary, "Data needs to be stored and organised in a way that will allow researchers to access, share, and analyse the material" (Tenopir et al., 2011, p.1).

6.4. Research data preservation practices

The conceptualisation of data curation models such as the CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008) was inspired by concerns about short and long-

term data preservation for access and re-use purposes. In this spirit, the study investigated research data preservation practices with a focus on the need for data preservation, storage facilities, data back-up strategies and research data infrastructure.

6.4.1. Need for research data preservation

Researchers commonly generate data in capacities of gigabytes, terabytes, petabytes, and eventually, exabytes (Walters & Skinner, 2011, p. 63). While researchers at UNI1 generally produced data in the capacities of between one GB and 100 GB, their counterparts at UNI2 produced data in the capacities of between 100 GB and one terabyte. A study by Ng'eno (2018) in Kenya reported the generation of huge amounts of data; research institutes generated data in capacities stretching to more than 500 petabytes. This difference can be attributed to the current study targeting individual researchers, while Ng'eno's (2018) study targeted research institutes operating at national level in a wide range of research areas including cereals, grain legumes, root, and tuber crops.

Although researchers generated data in large quantities, this valuable data had a limited lifespan because most researchers (99 or 52.9%) indicated that their data would only remain accessible for a period of between five and 10 years. This is regrettable because the two models informing this study envisaged that data should be preserved and accessible indefinitely. Malawian universities can draw inspiration from the University of Alberta Library which has embarked on an ambitious project that aims at preserving its digital intellectual capital including data so that it remains accessible for the next 500 years (Zvyagintseva, 2015).

All 187 (100%) researchers were of the view that it was scientifically sound to preserve the data they generated for re-use purposes. The study revealed that although both universities did not sponsor their researchers to conduct research, UNI2 researchers were funded by external organisations. In comparison, UNI2 researchers produced data in larger amounts than their UNI1 counterparts. The reason is that UNI2 researchers were involved in larger research projects funded by organisations or institutions while UNI1 staff commonly relied on self-funded research. This also explains the reason why UNI2 researchers generated data in capacities of up to one terabyte. The Director of Research at UNI2 commented that "For the college, I think it [research] is our number one earner of finances, so research is the one that keeps this college running". The health sector is one of the key priority areas of the Malawian Government (Malawi Growth Development Strategy (MGDS) (2017). In order to achieve effective health sector planning, the government has commited to making research driven

policies (MGDS, 2017). Noting that research is a linchpin for improved and sustainable health service delivery, Malawi's regional and international donor partners have opened funding taps to UNI2 for research purposes. Although UNI2 did not sponsor its researchers to conduct research, it rewarded them for publishing in creditable journals by paying a 100 USD per publication implying they were encouraged to conduct more research which translated into more research data production. On the contrary, there were no monetary rewards at UNI1 as commented by the Director of Research at UNI1 that "So, in the meantime, the only way we encourage researchers to publish is through promotion, that is, when academics publish, they get promoted". The fact that researchers at UNI2 conducted large research projects and produced more publications helps to explain the reason why there are more senior lecturers and professors at UNI2 than at UNI1.

6.4.2. Storage facilities

In the activities of *preservation planning and conceptualise* of the DCC Curation Lifecycle Model (Higgins, 2008), it is advised that researchers and curators need to plan well in advance the facilities that will be used to store and preserve the data. The results from library staff, researchers and directors of research confirmed that the key data storage facilities include personal laptops, office computers, external hard drives, flash disks and email accounts. In particular, all 187 (100%) researchers used personal computers. Findings showed a strong correlation between the two universities in the use of external hard drives and flash. Generally, personal computers, external hard drives and flash drives are readily available and their prices are within reach of most researchers, hence their widespread use by researchers in storing data. Universities in Malawi purchase digital storage facilities including computers, laptops, flash drives and external hard drives for use by their researchers and it is therefore expected that researchers use these facilities for data storage. The results replicate findings of prior studies in the USA and China. Most researchers in the USA and China use office computers, external hard drives, and flash drives in storing their research data (Chen & Wu, 2017; Schonfeld & Wulfson, 2013; Schumacher & VandeCreek, 2015).

However, unlike the current study, Schonfeld and Wulfson (2013) and Schumacher and VandeCreek (2015) found that researchers in the USA also used cloud-based applications. Failure to use these applications by researchers in the present study can be attributed to the lack of knowledge because, according to the findings, the universities offered email services (powered by Google) implying that they could easily use Google Drive, for example. Educating

and popularising the use of cloud-based services such as Google Drive and DropBox amongst researchers is of critical importance. This proposal is in agreement with the CCMF (Lyon et al., 2012) which indicates that community capability in data management can be enhanced by training its members in the relevant skills including tools and technologies (cloud computing). Twenty (55.6%) library staff indicated that although they had not helped researchers to save data files on a web-based or cloud application, they were ready to help. While researchers in Malawi had no access to a nation-wide infrastructure due to its unavailability, contrasting results were realised in France where Schöpfel, Ferrant, André, and Fabre (2018) found that 77% of researchers used nation-wide storage facilities and local servers.

6.4.3. Data backup strategies

Closely related to the aspect of data storage is the concept of data back-up. There is consensus in the literature that digital data is easily lost or corrupted due to software and hardware failure (Consultative Committee for Space Data Systems, 2002, p. 5; Cox & Pinfield, 2016). Most researchers (178 or 93.7%) had put in place some strategies to back up some of their research data. The common strategies used at both institutions include email accounts and external hard drives. The study revealed that researchers did not use local servers, central campus servers, web-based servers and data repositories. According to the results, there were no statistically significant differences in emails and external hard drives. As already noted, most researchers used free standing devices to store and back up their data; yet researchers lost their data through accidental damage of these storage facilities (see section 5.3.5.2 of Chapter Five). Hence use of free standing devices to store and back data implies that researchers will continue losing their data. Building central campus servers, web-based servers and data repositories could be the best strategy to strengthen the back-up of data. A study by Koopman (2015) noted that, drawing lessons from previous losses of data stored on free standing devices, researchers at UCT resorted to using cloud applications (Google Drive and Dropbox).

6.4.4. Research data infrastructure

The technical infrastructure supports research tools and services that are used at different stages of the research life cycle (Lyon, et al., 2011) which according to the DCC Curation Lifecycle Model (Higgins, 2008) include capture, appraisal, description, preservation, share, access and re-use. RDM will only become a success if a proper and robust data infrastructure is established (Shakeri, 2013; Tenopir et al., 2011). At UNI1, data infrastructure was non-existent. At UNI2, the college had established a data centre which operated on a small scale, hence it was neither

dependable nor sustainable. The Director of Research at UNI2 commented: "I cannot blame researchers for not using the facility [data centre] because we have limited capacity at the research support centre; we only have one data officer … and the data officer cannot cope". On the other hand, the Director of Research at UNI1 commented: "At the moment the answer is no, we don't have [data infrastructure]… researchers are only using office computers and their personal laptops". Absence of data infrastructure may have influenced most researchers (141 or 75.4%) to strongly or somewhat agree on the need for the universities to establish a process for short and long-term data management.

Data infrastructure extends far beyond the less reliable data storage facilities such as laptops and their peripherals such as flash drives and external hard drives. Dependable and robust data infrastructure is characterised by an integrated system incorporating hardware, software and human resources (Atkins, 2003). Brown, et al. (2015) advocate for a data management infrastructure system which facilitates general management of data including short and long-term preservation and access. In the present study, however, the absence of such infrastructure means that there is no formal system for collecting data from primary producers or data repositories; preparing data for long-term preservation; and assigning necessary metadata to facilitate access and re-use of data. Results suggest that there is an absence of university-furnished networked storage, non-networked devices; college or university-administered networks which, according to Schöpfel et al. (2018) and Schumacher and VandeCreek (2015), are indispensable for the successful short and long-term data management. Lack of such data infrastructure in the present study frustrates researchers from participating in RDM activities meaning the data they generate is not formally preserved. According to the CCMF (Lyon et al., 2012), investing in computer based large petabyte-scale research data storage is necessary for long-term storage, preservation, access and re-use of research data. In Europe, a study by the Permanent Access to the Records in Europe (2009) also revealed that poor and unreliable hardware, software, and support of computer environment scared researchers from embracing research data management activities. Observing a similar problem in South Africa, Denny et al. (2015) make a strong call to research grants organisations to seriously consider providing universities and research institutions with funding opportunities for developing and maintaining data curation infrastructure.

6.5. Competencies required for research data management

According to the CCMF (Lyon et al., 2011), human resource capabilities are inevitable for the successful accomplishment of all activities, roles and responsibilities in the data curation process. This section provides a discussion regarding the findings on competencies required by librarians and researchers to proficiently partake in RDM activities. In addition, the findings about the extent to which librarians offer RDM support to researchers are discussed.

6.5.1. RDM competencies required by librarians

According to Heidorn (2011), Newton et al. (2011) and Ray (2012), librarians are challenged to acquire new skill sets and competencies that will enable them to carry out new roles of digital curation. It is therefore not surprising that, in the present study, librarians needed various competencies and knowledge in RDM. Specifically, library staff at both institutions needed knowledge as proposed by the DCC Curation Lifecycle Model (Higgins, 2008) namely, curation standards and practices, models that guide data curation, long and short-term data curation activities, selection of data for preservation and data citation. There were no statistically significant difference between the two universities implying that the skills they needed were similar.

The findings suggest that the librarians needed these skills and competencies because they were incompetent in performing these activities. A number of factors could account for the librarians' skills gap in data curation. First, despite the emphasis on the need for training in digital curation (Charbonneau, 2013; Kim et al., 2013), the current study revealed that there were no training workshops organised to equip library staff with digital curation skills. In fact, 72.2% of library staff had not attended such workshops. In contrast to the findings of the present study, Kahn et al. (2014) report that in South Africa, LIASA had previously organised workshops on RDM which were facilitated by experts from the UK's DCC. According to Kahn et al. (2014), these workshops helped to instil librarians with basic and advanced skills in RDM. It is perhaps on this basis that Ng'eno (2018) suggests that library staff should be presented with opportunities to attend conferences on RDM where they can acquire the necessary skills and knowledge for managing their data.

Considering that workshops alone may not equip librarians with advanced skills and knowledge in RDM, Brown, et al. (2015), Higgins (2011), and Matlatse (2016) note that a widely held view is that LIS schools are better placed to offer skills in RDM specialities. Matlatse (2016) emphasises that LIS professionals should enrol with universities that offer

courses in RDM. Unfortunately, the only LIS school in Malawi at UNI1 does not offer any course or module in digital curation. Elsewhere, realising the knowledge gap in digital curation among librarians, the University of North Carolina at Chapel Hill in the USA offers a digital curation postgraduate certificate and similarly, a master's programme in digital curation is offered at Luleå University of Technology in Sweden (Higgins, 2011). In Africa, UCT has set the pace in RDM education becoming the first university in Southern Africa to offer a master's degree in data curation (Kahn et al., 2014, p. 299). The findings add weight to observations made by Latham (2017, p. 264) and Walters and Skinner (2011) that generally, there is the lack of skills on the part of information professionals to effectively support curation activities. Considering the momentum that the field of digital curation continues to garner from across research stakeholders, it is tempting to assume that more library schools will inevitably introduce courses in this area of specialisation.

On a positive note, librarians were competent in collecting data from creators, transferring preserved digital objects to repositories and providing access to stored digital objects to users. This is encouraging because Schmidt, Gemainholzer and Treloar (2016) highlight that librarians should possess knowledge and competencies in digital repositories, data discovery, data publication and the requirements of journal publishers. However, the absence of policies, as will be noted in section 6.6.2.5, implies that librarians were bound to perform poorly in these activities because the DCC Curation Lifecycle Model (Higgins, 2008) in its *ingest* action, stipulates that data selected for long term preservation can be transferred to an archive, repository, data centre or other strategic storage facilities by following documented institutional policies and legal requirements. More so, the absence of data infrastructure as evidenced in section 6.3.3.1, 6.3.3.2, 6.3.4.1, 6.4.2, 6.4.4, 6.6.1 and 6.6.2.1 implies that librarians were not putting their skills and knowledge into practice.

6.5.2. RDM competencies required by researchers

According to the DCC Curation Lifecycle Model (Higgins, 2008), RDM processes include capture, appraisal, description, preservation, share, access and re-use. To proficiently perform these activities, roles and responsibilities, the CCM Framework (Lyon et al., 2012), highlights the need for human resources capabilities. Data curation tools and knowledge or skills are hailed in the literature as necessary in achieving a meaningful data management process (Clement, Blau, Abbaspour, & Gandour-Rood, 2017; Curty, Crowston, Specht, Grant, & Dalton 2017; Houtkoop et al., 2018; Koltay, 2017; Van Horn & Gazzaniga, 2013). The present

study revealed that both researchers and library staff needed essential skills in RDM. Researchers at both universities needed knowledge and skills in preparing data management plans, short and long-term data preservation, identifying curation standards, metadata creation, migrating data to newer files and depositing data into archives. This means that at present, researchers in Malawian public universities lack these skills and competencies. The need for these RDM skills is not unique to the Malawian context because the literature shows that researchers in the developed world also need these RDM skills and knowledge. A study conducted in the USA by Schumacher and VandeCreek (2015) revealed that most professors did not possess the ideal knowledge and skills for managing their digital research data. The absence of workshops in RDM as revealed in this study could be mentioned as the key reason behind researchers' lack of knowledge and skills. The majority of researchers (141 or 74.2%) had not attended training workshops in RDM. Of the 46 (24.2%) researchers who attended the workshops, 35 said these workshops were organised by international organisations. Yet almost all staff involved with RDM will need training (Brown, et al., 2015). In that regard, a lesson can be drawn from Chinese researchers. According to Chen and Wu (2017), some Chinese researchers have gained these RDM competences through various mechanisms including special lectures, WeChat, online courses, phone/email, workshops and the library blogs.

6.5.3. RDM Support librarians offer to researchers

The researchers sought little or no assistance from the directors of research. Instead, they relied more on librarians and ICT experts and fellow researchers. Findings from the librarians confirmed that researchers approached them for help in various research activities especially in the areas of citation and referencing (30 or 83.3%), identification of journals (26 or 72.2%) and installation of data analysis software (26 or 72.2%). These findings echo observations by Walters and Skinner (2011) who note that librarians are key players in the research process. In the CCMF, Lyon et al. (2011) encourage RDM stakeholders to form alliances with various research stakeholders because collaboration adds value to the research process. Attaching this reasoning to the present study, the working relationship between librarians and researchers in preparing references and identifying creditable journals is a pointer to quality research.

However, librarians did not help researchers in RDM activities. This is regrettable because, according to Charbonneau (2013), Heidorn (2011), and Walters and Skinner (2011), researchers have poor knowledge or skills in RDM and their best hope is librarians. By nature of their job as knowledge managers, it is expected that librarians should offer regular and

demand training to improve researchers' RDM skills. The failure by librarians to help researchers in RDM in this study is attributed to the lack of knowledge in this area; the study confirmed that librarians lacked a compendium of skills in RDM. Taking a leaf out of the DCC Curation Lifecycle Model (Higgins, 2008), various prior studies (Charbonneau, 2013; Chen & Wu, 2017, p. 352; Kahn, et al., 2014; Walters & Skinner, 2011) and also considering that RDM is within the purview of library and information science, both universities need to invest in library staff RDM education to enable them to offer researchers help in data management plans, data discovery and retrieval, preservation and backup. If the current situation is not decisively reconsidered, researchers will develop a perception that librarians are unsupportive in their RDM activities.

6.6. Challenges in research data management

Despite the benefits that RDM brings to science, the research community is grappling with various challenges that continue to frustrate RDM efforts. This section discusses the challenges realised in the present study; it starts by discussing data losses amongst researchers before discussing the key challenges.

6.6.1. Data loss amongst researchers

Most researchers lost research data through various ways. From the three factors that were investigated, namely stolen storage facilities, accidental damage of storage facilities and obsolescence of technologies, the key contributing factor to data loss at both universities was accidental damage of storage facilities. Means, standards deviations and t-tests all showed a statistically significant differences across all dimensions implying that data loss was influenced by university affiliation. According to the findings, data loss was attributed to the lack of data storage infrastructure; most researchers stored their data on flash disks and personal laptops. This was emphasised by the Director of Research at UNI2 who commented that "…researchers keep most of their research data in their laptops and computers … and once these crash, the data is lost". In a study by Schumacher and VandeCreek (2015), it was also revealed that a greater number of researchers (74%) experienced data loss if their data was stored on free-standing devices (optical discs and external devices) thereby recommending that the use of an institution's network is the best secure option. Owing to their fragility and susceptibility to theft and losses, free standing devises (flash disks and personal laptops) can be described as high risk data storage facilities.

6.6.2. Challenges in RDM

There were a number of challenges that compounded data management activities at the two universities. The key challenges included lack of data infrastructure, lack of skills, lack of incentives and recognition, lack of collaboration in RDM activities and the absence of RDM policies. Both sets of quantitative results (researchers and library staff) showed that most challenges were not influenced by university affiliation because analysis of ANOVA revealed no statistical difference in most challenges.

6.6.2.1. Lack of infrastructure: Hardware, software and metadata standards

Lack of data storage and network infrastructure was cited by 117 (62.7%) researchers. Infrastructure in this context entails sustainable hardware and software (National Science Foundation, 2012; Permanent Access to the Records in Europe, 2009; Shakeri, 2013). Unavailability of the data infrastructure was attributed to lack of investments on the part of universities. The Director of Research at UNI2 commented that "Number one, I can say lack of investment, you need good servers and proper back up of power". Findings showed that while UNI2 had a data centre, UNI1 had none as confirmed by a statement from the Director of Research at UNI1 as follows: "At the moment the answer is no, we don't have [data infrastructure], but what I can say is that we have plans…". One form of data infrastructure which is important but was unavailable is software as cited by all 37 (100%) librarians and 115 (61.5%) researchers. Yet software is at the centre of integrating hardware, networks, data and users and in addition, it is fundamental in supporting current and future expected and unexpected needs (Kahn et al. 2014, p. 302; National Science Foundation, 2012, p. 4).

Schumacher and VandeCreek (2015) report that loss of digital objects including data is largely attributed to incompatible hardware or software. In addition, the DCC Curation Lifecycle Model (Higgins, 2008) emphasises the need for acquiring and constantly updating hardware or software which ensures continued access to data files on a long-term basis. These results are not so different from those reported by Ng'eno (2018) who noted that statistical software such as SPSS and Genstart were the most common software in Kenyan research institutes implying that there was an absence of software for supporting RDM activities. The two Malawian universities can learn from the de.NBI-SysBio in Germany which uses the SEEK software to implement its four principles of *Findability*, *Accessibility*, *Interoperability*, and *Reusability* (Wittig et al., 2017, p. 229) which conform to the DCC Curation Lifecycle Model (Higgins, 2008), namely data capture, appraisal, description, preservation, share, access and re-use.

Constructing a robust and dependable data infrastructure has never been inexpensive. Being public universities operating on tax payers' money with meagre and unsustainable budgets implies that the little resources are channelled towards critical university operations such as salaries and other day-to-day activities. The failure by the universities to invest in data infrastructure could be the reason most researchers (147 or 78.6%) and library staff (29 or 78.4%) indicated that they lacked university support in their RDM activities. For UNI2 to have a data centre is somewhat reasonable because according to UNI2's Director of Research, in addition to financial support from the government, "it's [research] our number one earner of finances, so research is the one that keeps this college running" meaning the college was able to construct the data centre with financial support from external funders. However, the data centre at UNI2 is not sustainable as it operates on a small scale prompting most researchers to keep data in individual departmental laboratories with the result that most data is not linked to the data centre. Similarly, Ng'eno (2018) noted that despite research institutes in Kenya generating large amount of research data, access, sharing and re-use of this data was problematic because it was scattered among different research institutes, departments and individual researchers. A similar problem was observed in the UK by Brown, et al. (2015) where existing data infrastructure is distributed across different faculties or multiple sites. The danger of such an arrangement, according to Brown, et al. (2015), is that it becomes problematic in coming up with a unified storage solution or centralised data management centre. This is because implementing standalone data storage systems can lead to the use of dissimilar or incompatible metadata standards which eventually pose challenges to the interoperability of the systems. Essentially, the DCC Curation Lifecycle Model (Higgins, 2008) proposes that data selected for preservation should be stored in a manner that adheres to relevant standards. It is therefore understandable to note that in the present study, some researchers mentioned a lack of metadata standards as one of the challenges affecting their RDM activities.

Challenges identified in the present study are similar to those reported elsewhere such as in China and Europe. Chen and Wu (2017) report that undependable data storage infrastructure was a key challenge that research institutions needed to address in order to realise practical data management. In Europe, it was noted that a lack of sustainable hardware, software, and support of the computer environment is the most important threats to digital data preservation (Permanent Access to the Records in Europe, 2009).

6.6.2.2. Lack of skills: Data literacy, metadata creation and standards

Skills and training are highlighted by the CCMF (Lyon et al., 2011) as necessary for data management. The skills and knowledge are key in performing various activities and actions involved in the data curation lifecycle as proposed in the DCC Curation Lifecycle Model (Higgins, 2008). The researchers lacked general data management skills, in particular data sharing (see section 6.3.3.3), tracking updates to data and metadata creation and its standards. The aspect of difficulty in finding data produced by other researchers can be linked to two possible reasons. The first reason could be that researchers do not have adequate data information literary skills which are critical for searching, retrieving, evaluating and using data. The majority of data sets are stored in online or offline databases but locating them can be a challenge to most researchers. Prado and Marzal (2013) state that researchers need to be equipped with data literacy skills that can enable them to search, retrieve, and use data to solve their problems.

Owing to the fact that librarians are well grounded in information literature delivery, they are better placed to impart data information literacy skills to researchers. Perhaps even more troubling, however, the findings of the study show that library staff lacked most skills in RDM. Librarians had basic digital preservation management skills such as helping to save files on a disk, USB drive and computer hard drive as cited by 35 or 97.2%. This is perhaps the reason why some authors have suggested the need for data information literacy for both librarians and researchers (Koltay, 2017) especially considering that digital curation is an emerging concept (Higgins, 2011). The second reason for the difficulty in finding re-usable data could be attributed to most researchers not sharing their own data. This could well be the reason that the majority of library staff (35 or 94.6%) were not able to access data for curation purposes. It was therefore, not surprising that library staff did not find the aspect of larger amounts of data to curate a challenge meaning there was actually no data for them to curate. The failure by researchers in the current study and elsewhere to share their data and their reasons were documented in section 6.3.3.3.

Metadata is crucial for the discoverability of data (Bryant et al., 2017) because they provide important information for data re-use and further preservation (Costello, 2009; Enke et al., 2012; Scot, 2014; Tenopir et al., 2011; Yoon, 2015, p. 173; Yoon & Schultz, 2017). However, nearly all (182 or 97.3%) researchers in the present study lacked knowledge in creating or assigning metadata to their data. It is therefore not surprising that 173 (91.1%) researchers did

not assign metadata to their research data implying that only a very small number of researches with a score of 14 (7.4%) could assign some form of metadata to their research data. The lack of knowledge and skills can be attributed to the failure by universities to organise workshops on RDM. These results are not unique to the present study because Koopman (2015) reports that some researchers at UCT in South Africa, did not assign any metadata to their data attributing the results to a lack of knowledge in metadata creation. The unfortunate part is that, in the present study, the library staff who by nature of their job are supposed to possess metadata skills, lacked these skills. Librarians need to acquire knowledge for working with various forms of metadata standards such as MARC, Dublin Core, METS, MODS and PREMIS (Kim et al., 2013, p. 74). Complications arising from the failure to adopt common metadata and standards are further discussed in sections 6.3.3.3, 6.3.4.1 and 6.6.2.1.

6.6.2.3. Lack of incentives and recognition

In the CCMF, Lyon et al. (2011, p. 44) argue that "intensive research is most likely to flourish in communities where data is valued highly: where researchers are recognised for their data contributions, and high standards are expected of data entering the research record". Most UNI1 (89 or 86.4%) and some UNI2 (35 or 41.7%) researchers did not partake in RDM activities because they were not incentivised. As Bryant et al. (2017) and Scott (2014) warn, building an imposing data infrastructure alone does not spontaneously equate to its anticipated use. Hence, researchers need to be rewarded for their engagement in data management activities. As discussed in sections 6.3.3.3, 6.3.4.1 and 6.3.4, data management activities such as data sharing comes with no inherent rewards to researchers. At UNI2, 37 (41.7%) researchers had some motivation, perhaps because most of the research is funded by international organisations which require researchers to clearly stipulate data management plans as part of the criteria for awarding the grants. The only known factors that are key in compelling researchers to engage in data management are funders and publishers (Brown, et al., 2015; Huang et al., 2012).

Considering that RDM initiatives are hard to sell, some researchers have proposed the need to introduce better incentives. It is argued that it is essential to provide enough funds for RDM activities and to formulate, adopt and enforce policies that maximise the citation of data sets (Brown, et al., 2015). This is necessary because, in their CCMF, Lyon et al. (2011) argue that one of the most successful rewards is where all contributions by researchers are recognised and rewarded, through established procedures and measures. Recent developments are likely to

accelerate data sharing through repositories. Pitt and Tang (2013) report that digital object identifiers (DOIs) for data are now being assigned implying that DOI for data sets has become an ISO standard. The implication is that consistent permanent links to data sets are now provided. Another breakthrough is that Thomson Reuters has created a Data Citation Index which track data re-use (Pitt & Tang, 2013) thereby providing appropriate recognition to those who collect and share data. However, the literature remains mute as to whether universities should take into account the publication of data when appraising their staff.

6.6.2.4. Lack of collaboration in RDM

The CCMF (Lyon et al., 2011) underlines the importance of collaboration in RDM which can occur within the discipline, across disciplines, across sectors and the public. In the present study, the challenge was that although researchers consulted librarians in some research activities, they hardly engaged them in RDM activities. Put differently, researchers and librarians did not involve each other in data management activities. It is therefore not surprising that 35 (94.6%) librarians said the key challenge was that researchers did not engage them in data management activities. However, it is the view of this study that library staff should not wait for researchers to come to them; rather it is their responsibility to take their services to researchers. Considering that researchers were able to seek out other research services, it is possible to speculate that they may as well have sought assistance on RDM but librarians did not provide these services because they lacked skills and knowledge in RDM. Specifically, although recommended by the DCC Curation Lifecycle Model (Higgins, 2008), library staff could not help researchers decide which data is important to preserve - *appraise and select* (26 or 72.2%); decide which data can be safely shared - *access, use and reuse* (30 or 83.4%); determine standards for identifying sensitive data - *community watch and participation* (21 or 58.3%); assign metadata - *description and representation* (17 or 47.2%) and determine data storage and long-term preservation - *ingest* (22 or 61.1%).

The DCC Curation Lifecycle Model (Higgins, 2008) highlights that curators (librarians in this context) need to collaborate closely with the data providers to understand the data and assign the best descriptors. It is recommended that collaboration should begin at the *conceptualise* stage of the DCC Curation Lifecycle Model (Higgins, 2008) or at the research development or proposal writing stage so that librarians can help researchers articulate in their research proposals the methods and procedures for long-term preservation of the research data (Heidorn, 2011).

6.6.2.5. Absence of data management policies

Bryant et al. (2017) warn that constructing a data infrastructure alone is not enough if adequate data management policies are not developed, implemented and enforced. According to the CCMF (Lyon et al., 2011), legal frameworks are necessary for RDM; these frameworks should be clear, articulate, consistent and responsive to all stages and activities of data curation lifecycle. Indeed, lack of data management polices was cited by 30 (83.3%) librarians, 127 (67.9%) researchers and both directors of research as a contributing factor to poor data management practices. For instance, the Director of Research at UNI1 commented that "So, maybe let's start by pointing out that at the moment, we do not have a clear policy and here we are talking about ownership of research data". Although universities were committed to formulating RDM policies, they were challenged with uncertainties regarding their capability in coming up with policies that could coherently harmonise the various components of RDM, namely preservation, sharing, re-use of data and reward systems. To avoid formulating controversial, unfavourable or unresponsive RDM policies, Denny et al. (2015) suggest the need to take into account the views and values of the concerned research stakeholders in a particular context.

The absence of data management policies was likely to breed various undesirable implications in RDM activities at the two universities. First, researchers, library staff and university stakeholders were at a liberty to not take data management seriously as there was no policy compelling them to do so. Secondly, the university authorities could not appropriately plan and implement data sharing activities considering that there was no policy to compel them; they could express commitment but implementation might be problematic. In other words, policies and regulations did not exist as envisaged by the DCC Curation Lifecycle Model (Higgins, 2008) which stipulate that data curation actions should be executed in accordance with established institutional polices. In line with the CCMF (Lyon et al., 2011), the absence of RDM policies suggests that RDM activities at the two universities were conducted haphazardly or in uncoordinated manner. Anane-Sarpong et al. (2017) also note that a lack of harmonised guidelines and structures are responsible for the slow uptake of RDM activities in Uganda and Tanzania. The impact of policy frameworks in this study is different from prior studies. While the absence of policies led to poor data management practices in the present study, an international study by Huang et al. (2012) revealed that, although research institutions had enacted policies, some of the polices were 'discriminatory' because they favoured certain data sets over others.

6.7. Summary of chapter

The chapter discussed and interpreted the results which were presented in Chapter Five. The discussion was shaped by the four research objectives that were developed in Chapter One to help address the research problem. In addition to contextualising the findings in prior related literature, theoretical frameworks - CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008) which underpinned the study were used to provide meaning to the findings. In the course of discussion, efforts were made to show differences and similarities of the findings of the present study to those of related prior studies.

The study established that the two universities were involved in research which was valued highly as it helped to position and expose the universities in the international higher education landscape. The key data formats included digital texts, audio recordings, spreadsheets and specimens. In addition to sharing their own data, researchers were willing to use data previously produced by other researchers and research institutions. However, difficulty in findings data for re-use frustrated data re-use activities. It was revealed that most researchers were involved in data preservation and back-up activities. The key data storage and preservation facilities included personal computers, office computers external hard drives, flash disks and email accounts; the same facilities were used for data back-ups. The universities had not invested in campus-based data infrastructure hence researchers were deprived of reliable and robust data storage facilities. The study demonstrated that researchers and librarians had knowledge gaps in most data management activities. Competencies and skills required included preparing data management plans, short and long-term data preservation, identifying curation standards, metadata creation, migrating data to newer files and depositing data into archives. The study exposed various factors that stymied RDM activities; they included lack of data infrastructure, lack of skills, lack of incentives and recognition, lack of collaboration in RDM activities and absence of RDM polices.

The next chapter (Chapter Seven) focuses on summarising the study findings; making conclusions and recommendations; highlighting contribution of the study and discussing the proposed model for RDM.

CHAPTER SEVEN

SUMMARY, CONCLUSION AND RECOMMENDATIONS

7.1. Introduction

The chapter focuses on presenting the summary of the findings, conclusion of the findings, recommendations and contribution and originality of the study. As an extension to a demonstration on the contribution and originality of the study, the chapter presents and discusses the proposed model for RDM in Malawi. The chapter concludes by making suggestions for further research which is necessary to fill the research gaps which were not addressed by the present and other prior studies.

The study explored RDM practices in public universities in Malawi with the intention to identify best practices for either improving or implementing RDM. The study was inspired by the fact that most researchers in public universities in Malawi were already involved in research production as evidenced by the growing number of research papers published in both open access and subscription based journals. Considering that data is hailed as an impetus for current and future discoveries through its re-use, it was necessary to understand how researchers stored, preserved, shared, accessed and re-used this data. Four research objectives were formulated to guide the study: to determine research data creation, sharing and re-use practices in public universities in Malawi; to investigate research data preservation practices in public universities in Malawi; to investigate competencies that librarians and researchers needed to effectively manage research data in public universities in Malawi; and to find out the challenges that affected the management of research data in public universities in Malawi.

The study was underpinned by the pragmatic paradigm and was further guided by two theoretical models namely, the CCMF (Lyon et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008). The study population included researchers, librarians and directors of research. While questionnaires were used to collect data from researchers and librarians, an interview protocol guided the interviews with directors of research. The SPSS was used to capture quantitative data and to perform percentages and frequencies. An independent sample t-test and ANOVA in SPSS statistical package were used to analyse the data. Qualitative data was analysed using thematic analysis (Braun & Clarke, 2006).

This chapter is divided into three key sections, namely summary of findings, conclusion, and recommendations. In addition, the chapter discusses the originality of the study, presents a proposed model for RDM and makes suggestions for areas for further research.

7.2. Summary and conclusion of the findings

This section summarises the key concepts arising from a discussion conducted in Chapter Six. In addition to demographic profiles of the respondents, issues highlighted in this section include research data creation, sharing and re-use practices (research output, data format(s), data sharing practices and data re-use practices); research data preservation practices (the need for data preservation, storage facilities, data back-up strategies and research data infrastructure); RDM competencies (RDM competencies for librarians and RDM competencies for researchers and support of RDM librarians offer to researchers); and challenges that affect RDM (data loss amongst researchers and challenges in RDM).

7.2.1. Summary of the demographic profiles of the respondents

The study established that 16 (44.4%) library staff were from UNI2 and 20 (55.6%) were from UNI1; 103 (55.1%) researchers were from UNI1 and 84 (44.9%) from UNI2. The study examined various demographic variables of participants, particularly, gender, qualifications, rank and university affiliation. The study established that there were more males than females. For library staff, there were 25 (69.4%) males and six (16.9%) females while five (13.9%) did not indicate their gender. For researchers, 135 (72.2%) were males and 52 (27.8%) were females. In terms of qualifications, the study established that for library staff, the majority had a diploma with a score of 10 (27.8%) followed by those with an LIS Certificate (eight or 22.2 %), Bachelor's Degree (eight or 22.2 %), Master's Degree (seven or 19.4%) and only one had a PhD. Two library staff respondents did not indicate their qualification. For the researchers, 101 (54%) had a Master's Degree, 77 (41.2%) had a PhD, eight had a post-PhD and one did not respond. The study revealed that the majority of library staff held the rank of senior library assistant (15 or 41.7%) followed by the rank of library assistant (10 or 27.8%), assistant librarian (eight or 22.2%), senior assistant librarian (two or 5.6%) and one (2.8%) at the rank of university or college librarian. For the researchers, the majority were at the rank of lecturer (103 or 55.1%) followed by senior lecturer (58 or 31%), associate professor (15 or 8 %), professor (10 or 5.3%) and one (.5%) professor.

7.2.2. Research data creation, sharing and re-use practices

The first objective of the study addressed issues of research data creation, sharing and re-use practices in public universities in Malawi. In terms of data creation, the study established that both universities were involved in intensive research activities. The study established that research was highly valued by the universities for three key reasons: it helped to expose and position the universities in the international education landscape; it helped to bring funds to the universities; and it was a driving engine for the national economy. Many papers originating from these research outputs were published in creditable journals which were recommended to researchers by librarians. The study established that by sharing research findings in journals and other publishing outlets, the researchers were contributing to the advancement of science which is a driving engine of the national and world economies. The results were consonant with the CCMF (Lyon et al., 2011) which underlines that open communication of research methods and results contributes to scientific progress.

The study established that the most common data formats were digital texts, audio recordings and spreadsheets; the less common or non-existent types of data forms included spatial data, computer codes and video recordings. The deluge of data in digital form was attributed to advancements in ICTs which have decisively penetrated the research community thereby impacting almost each and every aspect of the research activity. In particular, the study noted that ICTs such as computers and related technologies have escalated the rate of research production. The study established that audio recordings were popular at both universities whereas specimens were only popular at UNI2; the reason was that while interviews are popular in the social sciences and humanities; and health sciences, specimens are mostly common in clinical research where specimens are generated from samples legally and ethically extracted from human beings. Ethical complications associated with data collection through video recordings were contemplated to be the reasons researchers shunned generating data in video format.

The study ascertained that the majority of researchers shared or were willing to share the data they generated. Two factors influenced researchers to share their data and they included personal initiatives and journal policies. On the other hand, research funders, university policies and open access initiatives did not compel researchers to share their research data. The study noted that journal publishers played a key role because most researchers were publishing in creditable journals which also demanded researchers adhere to data sharing policies adopted

by those publishers. The only drawback, however, was that most journal policies were rather weak and lacking enforcement by the publishers. Research funders which are hailed as key in compelling researchers had no influence in compelling researchers to share the data. Most researchers at UNI1 conducted self-funded research and it was not possible for such policies to play a key role in this regard; at UNI2, despite conducting donor funded research, the impact of research funders' policies was negligible as most research grant organisations failed to enforce these polices. The study noted that poor data sharing was partly attributed to the government owned National Commission for Science and Technology which was failing to fulfil its mandate of sponsoring research in universities due to financial challenges.

The findings showed that data sharing tools used included external hard drives and emails. Social networks, blogs, wikis, journal websites, funder's websites, university websites, clouds, principal investigator's website, national network, and regional network were not used for data sharing. The study noted that while failure to use web-based applications such as Google Drive to share data was attributed to the lack of knowledge and skills, failure to use a national network was attributed to its unavailability in Malawi.

The study ascertained that although researchers were willing to share data, some were not due to lack of data sharing skills, lack of metadata and their standards. The concept of data sharing was new in Malawi and failure by universities to conduct workshops on the same exacerbated the problem. The study also noted that researchers found it impractical to share their data because it lacked proper description (metadata) and standards or guidelines, hence it was not widely readable. Failure to document data was also attributed to the lack of skills and knowledge. Lack of time was another factor that restricted researchers from sharing their data; researchers were equally busy with other pressing roles of teaching, supervising students, conducting research and participating in community services.

In order to strengthen data sharing at the two universities, two conditions were established: there was a need to allow researchers to place restrictions on the data they shared and compel re-users to cite the data. By placing access restrictions on the data they shared, researchers were aiming at countering possible cases of misuse or misappropriation of their data by re-users. Considering that sharing data comes with no inherent rewards, the only way researchers could get recognition was through having their data cited by re-users.

Most researchers used or were willing to use data generated by other researchers or research institutions. The study established that data re-use came along with some benefits including

advancements in science, affordable costs associated with data re-use and convenience on the part of data re-users with regard to issues of ethics clearance. It was established that conducting fresh research was more costly than re-using available data and data re-use optimised the use of resources. The study established further that re-use of data was encouraged because researchers, particularly students, struggle to obtain permission from authorities or organisations to conduct studies and data re-use removed this burden. It was established that data re-use enhanced science through re-analysis of available data thereby arriving at new breakthroughs.

The study established that although researchers were involved or willing to share data, two interrelated factors frustrated their efforts in re-using data; difficulty in accessing re-usable data and lack of metadata (metadata standards). The inability to access re-usable data was attributed to the absence of metadata; re-users could not understand the context in which the data was created, the purpose of the data, creators of the data and legal aspects associated with access and re-use of the data. However, the study established that the issue of trust did not affect researchers' intentions to re-use data; they had trust in data produced by other researchers and would not misinterpret it as long as it was accompanied by comprehensive metadata.

7.2.3. Research data preservation practices

The second research objective sought to determine data preservation practices with a focus on the need for data preservation, storage facilities, data back-up strategies and research data infrastructure.

The researchers at UNI1 generated data in capacities of between one GB and 100 GB while UNI2 researchers produced data of between 100 GB and one terabyte. All the researchers were of the opinion that data should be preserved for data re-use purposes. However, it was established that the data had a limited lifespan because it could no longer be accessible after a period of 10 years.

The key data storage facilities included personal laptops, office computers, external hard drives, flash disks and email accounts. All researchers used personal computers. The use of external hard drives and flash drives was attributed to the availability of these tools purchased by universities for use by their staff. There was no national wide infrastructure for researchers to use for their data storage and preservation activities.

The study ascertained that researchers had put in place some strategies to protect their data from loss. The key strategies included email accounts and external hard drives. However, local servers, central campus servers, web based servers and data repositories were not used because they were not available. It was established that continued use of external hard drives would escalate data loss.

UNI1 had no data infrastructure while UNI2 had constructed a data centre which was, however, operating on a limited scale. There was an absence of data infrastructure such as university-furnished networked storage and college or university-administered networks. There was no formal system for collecting data from primary producers or data repositories; preparing data for long-term preservation; and assigning necessary metadata to facilitate access and re-use of data.

7.2.4. Competencies required for research data management

The third objective of the study examined competencies that librarians and researchers needed to effectively manage research data.

The findings of the study revealed that the two universities did not have adequate skills in RDM. For library staff, they needed skills and knowledge in various RDM activities including curation standards and practices, models that guide data curation, long and short-term data curation activities, selection of data for preservation and data citation. These skills and competencies were missing amongst library staff due to the absence of RDM workshops and formal education in RDM. Impressively, library staff were competent in collecting digital objects from creators, transferring preserved digital objects to repositories and providing access to stored digital objects. Despite having some skills in the RDM specialities mentioned, library staff did not put into practice their skills because there were no formal data management activities in the universities. Researchers needed various RDM skills that included preparing data management plans, short and long-term data preservation, identifying curation standards, metadata creation, migrating data to newer files and depositing data into archives. The absence of workshops in RDM was attributed to a lack of such skills and knowledge; universities did not organise RDM workshops.

The librarians did not help researchers in RDM activities. Instead, it was established that librarians offered general research support services particularly, citation and referencing, identification of journals and installation of data analysis software. Failure by librarians to offer RDM services was attributed to the lack of skills and knowledge in various specialties of RDM.

7.2.5. Challenges affecting research data management

Most researchers lost research data through accidental damage of storage facilities. The study noted that continued use of self-standing devices, which were unreliable and prone to damage, meant researchers would continue losing their research data. Universities in Malawi are confronted with RDM challenges such as: a lack of data infrastructure, lack of skills, lack of incentives and recognition, lack of collaboration in RDM activities and absence of RDM polices. The lack of infrastructure was due to a lack of investment by the two universities in hardware, software and network infrastructure. Although UNI2 had a data centre, the facility operated on a small scale and could not support large data sets generated by researchers.

7.3. Conclusion

Lunenburg and Irby (2008, p. 240) and Williamson and Bow (2000, p. 281) provide three useful points to be borne in mind by postgraduate students when writing conclusions and recommendations of their theses: conclusions and recommendations should clearly relate to the findings; the researcher should not over-conclude - sweeping conclusions and generalisations should be avoided; and the researchers should show whether the objectives or questions have been addressed or not. Further lessons about writing conclusions are drawn from Bouma and Atkinson (1995, p. 227): writing a conclusion involves re-stating the key results of the research study and drawing the implications on the research objectives at hand. In summary, Lunenburg and Irby (2008, p. 240) characterise conclusions as "assertions based on the findings".

This section provides conclusions synthesised from the findings from each research objective identified in Chapter One and restated in section 7.1. Conclusions are synthesised from the discussion of the findings presented in Chapter Six.

7.3.1. Research data creation, sharing and re-use practices

Public universities in Malawi are involved in research activities thereby generating substantial amount of research data. Most data are generated in formats of digital texts, audio recordings and spreadsheets. An ethical dilemma affected the generation of data in the form of video recordings. The involvement of universities in research activities was inevitable because of the many benefits research brought to the university environment; exposing the universities to the world with the aim to attract esteemed staff and exceptional students; and attraction of funds for running the universities influenced universities to value research highly. Malawian universities had put in place some mechanisms for rewarding researchers involved in research

publications; such rewards included promoting researchers to senior positions such as senior lecturer and other higher ranks within the university structure. UNI1 was encouraged to learn from UNI2 in improving its reward system by incorporating monetary rewards in its reward mechanism. The failure by the universities to sponsor researchers in their research activities was a point of concern especially at UNI1 where most researchers conducted self-funded research. The study concluded that putting in place the reward system that boosted research production and data sharing was necessary because this could contribute to scientific progress through data re-use.

Owing to its potential in fuelling science progress, debates about data sharing are increasingly dominating the international research agenda. However, the concept is not yet common practice in academia. Universities in Malawi are slowly but surely embracing the concept as evidenced by the higher number of researchers who were willing to share their data. The role of journal publishers in enforcing Malawian researchers to share their data was key. This was in addition to personal initiative by researchers to have their data accessible to others; this was the right direction towards the attainment of free data sharing in Malawi. Although researchers were commended for their willingness to share data, the data sharing tools such as free-standing devices that they used were inappropriate and undependable. Emails were secure but the problem was that not everyone could access data shared via emails; access to research data is key in the attainment of meaningful scientific progress. UNI2 had set the pace in RDM by establishing a data centre whose activities were however limited and unsustainable. Tools such as social networks, blogs, wikis, journal websites, funder's websites, university websites, clouds, principal investigator's website, national network, and regional network were not utilised because of various factors which need to be addressed. To encourage researchers to share data which comes with no inherent rewards, it was necessary to accord them some powers over the data they shared; this was in addition to ensuring that re-users of the data were encouraged to cite it as part of recognising those who shared the data. Due to absence of rewards, researchers opted not to allot their already limited time to data sharing activities. Data sharing is a new concept in Malawian universities. Therefore, lack of skills and knowledge is a key challenge. The problem is aggravated by the failure of universities to organise workshops on RDM. Lack of these skills meant that researchers could not make any attempts to document their data, hence it was impossible for them to share such data as it was not readable and not ready for public consumption.

Data sharing alone does not translate into data re-use. The majority of researchers were willing to use data produced in previous studies. Data re-use was hailed by directors of research as a driving force for further science discovery through its re-analysis. It was further hailed as a source of solace to struggles researchers particularly students go through in their research ethics clearance processes. Conducting research is naturally expensive and data re-use was praised for optimising resources in the research process. Regardless of the potential benefits accrued to data re-use, factors that frustrated researchers to re-use existing data included difficulty in accessing re-usable data and lack of metadata and their corresponding standards. On the one hand, it was encouraging that researchers had trust in data produced by others; they had adequate skills to re-analyse secondary data; and more significantly, they would not misuse such data. The interconnectedness of difficulty in finding reusable data and absence of metadata was discernible; re-users of data found it difficult to find such data because it was not properly documented. Comprehensive and accurate metadata are therefore vital in informing re-users about variables represented by the data.

7.3.2. Research data preservation practices

Researchers generated data in larger amounts reaching up to one terabyte. Researchers at UNI2 produced data in larger capacities than their counterparts at UNI1. Preservation of this data was key in ensuring current and future data re-use. Both universities acknowledged that it was morally right to preserve data. Unfortunately, this valuable data could only be accessible in the next five to 10 years due to poor data preservation strategies. The state of data storage facilities was not at par with conventional standards. Researchers relied on personal laptops, office computers, external hard drives, flash disks and email accounts. Overall, personal computers were the most prevalent. Provision of these data storage facilities by universities seemed to suggest that their use by researchers was influenced by their availability. However, web based tools such as the clouds were not utilised attributing their non-use to lack of knowledge and skills. Failure by the Malawi government to construct a national wide data infrastructure was the reason researchers could not use such a facility. A natural extension of data storage aspect is data back-up. Noting that loss of data was inevitable, most researchers had adopted some basic data back-up strategies. Facilities used by researchers to store data (email accounts and external hard drives) were also used for data back-ups. Again, their use was influenced by their availability. Mainstream back-up facilities namely, local servers, central campus servers, web based servers and data repositories were not utilised by researchers.

Poor storage and back-up facilities confirmed absence of data infrastructure. However, a thin line could be drawn between UNI1 and UNI2 in their data infrastructure initiatives; while UNI2 had a small and unstainable data centre, UNI1 had none. What was regarded as dependable infrastructure by researchers were paradoxically less reliable data storage facilities namely, laptops and their peripherals such as flash drives and external hard drives. These tools posed high risks to data through accidental damage and misplacement or theft. Robust and dependable infrastructure (university-furnished networked, college or university-administered networks) which is characterised as integrated system of hardware, software and human resources necessary for short and long-term data preservation was unavailable. Based on the absence of such data infrastructure, it was concluded that researchers could continue losing their data and that there was no formal process for collecting, preserving, accessing and reusing data.

7.3.3. Competencies required for research data management

Researchers and librarians acknowledged that they needed various skills and knowledge in various RDM activities. The skills which are important in the fulfilment of RDM activities but were lacking in librarians and researchers included preparing data management plans, short and long-term data preservation, identifying curation standards, metadata creation, migrating data to newer files and depositing data into archives. Absence of skills and competencies were attributed to unavailability of training workshops in addition to the fact that RDM was a new concept in Malawian universities. Formal education in Malawi could not offer formal RDM training to librarians because the only library school in Malawi was yet to introduce courses in this emerging speciality. It could be concluded that inevitably, librarians who are by nature knowledge managers, needed to acquire these skills so that they could in turn either adequately serve researchers' needs or help train these researchers in RDM. It turned out that due to lack of skills on the part of librarians in RDM, researchers only helped researchers in general research areas save for RDM. Ignoring the red flag that if little is done by librarians to acquire skills that will enable them adequately serve RDM interests of researcher is counterproductive; the value of librarians will be eroded in the minds of researchers as natural RDM experts.

7.3.4. Challenges affecting research data management

Universities in Malawi were faced with various RDM challenges such as: lack of data infrastructure; lack of RDM skills; lack of incentives and recognition; lack of collaboration in RDM activities; and absence of RDM polices. Due to some of these challenges, most researchers experienced data loss through accidental loss of data storage facilities which were

mostly free standing devices. It can be concluded that lack of data infrastructure was due to lack of investment in this aspect. On the lack of incentives and recognition, it was concluded that RDM comes with no inherent benefits hence, researchers were reluctant to direct their efforts and time in RDM. The study concluded that lack of collaboration in RDM activities was attributed to the lack of skills and knowledge by librarians in RDM hence could not handle RDM queries from researchers. It was concluded further that absence of RDM policies adversely affected RDM activities: librarians, researchers and university management were not formally compelled to partake in RDM activities.

7.4. Recommendations

The study has presented evidence that there were a number of challenges confronting RDM activities in Malawian public universities. To mitigate these challenges, this section presents some recommendations that could reinforce and smoothen RDM implementation in Malawian public universities. The recommendations are presented in line with the findings emanating from each research objective.

7.4.1. Research data creation, sharing and re-use practices

The study established that the universities had put in place some mechanisms to encourage research production amongst lecturers. These reward mechanisms included promoting researchers to senior positions within the university system. UNI2 took a step further by rewarding researchers with USD 100 for any research paper published in creditable journals. The study recommends that UNI1 should introduce monetary rewards for its researchers to encourage them to conduct more research and to share the data which translates into scientific progress through data re-analysis. It was further noted that UNI1 researchers conducted self-sponsored research. The study recommends that UNI1 should either set aside a research budget or find mechanisms to attract external funders for research activities.

Data sharing was restricted to the use of less dependable free-standing devices and emails. The study recommends that universities should popularise data sharing using social networks, blogs, wikis, and clouds which are free web-based applications. The study recommends further that universities should collaborate with key research stakeholders including donors and government to come up with a blueprint for establishing campus wide and national networks for sharing data. Since researchers allocated little or no time to data sharing due to the lack of time, the study recommends that universities should reward researchers who share their data. Data sharing and re-use were affected by the lack of metadata which was ascribed to a lack of

skills on the part of researchers and librarians. It is therefore recommended that universities should organise regular or demand driven training to equip researchers and librarians with data sharing skills and knowledge.

7.4.2. Research data preservation practices

The generation of data in universities was high. Data was stored using undependable devices which included personal laptops, office computers, external hard drives, flash disks and email accounts; the same facilities were used to back-up data. UNI1 had no infrastructure for data management; UNI2 had a data centre which apart from being unsustainable, was not made known to researchers. It is therefore recommended that the universities should construct a reliable infrastructure that researchers with the help of librarians can use to capture, store and preserve research data. In the meantime, the study recommends that UNI2's Director of Research should work in collaboration with the library to market the services offered at its data centre to researchers. It is further recommended that librarians should take a leading role in training researchers in the use of web-based applications (Google Drive and DropBox) to store and preserve their data.

7.4.3. Competencies required for research data management

The study revealed an acute lack of RDM competencies and skills amongst librarians and researchers. Insufficiency in RDM skills and knowledge was as a result of the failure by universities to conduct RDM workshops and the absence of formal education in RDM in Malawi. The study recommends that universities should conduct training that can adequately equip researchers and librarians with RDM skills. Considering that failure by researchers to share data via open access data repositories was attributed to a lack of knowledge, it is recommended that workshops on RDM should also cover open access concepts that relate to data sharing. Experts in RDM such as those from UK's Digital Curation Centre who have offered similar services in South Africa and elsewhere, could be invited to facilitate such training. It is recommended further that the LIS department in Malawi at UNI1 should revise its curricula to incorporate courses on RDM; this should be in addition to offering periodic RDM workshops to librarians and researchers in Malawi.

7.4.4. Challenges affecting research data management

RDM efforts were confronted with many challenges that included lack of data infrastructure; lack of RDM skills; lack of incentives and recognition; lack of collaboration in RDM activities;

and absence of RDM policies. Recommendations regarding the lack of data infrastructure; lack of RDM skills and lack of incentives and recognition are already presented in previous sections. The lack of collaboration between researchers and librarians thwarted RDM initiatives. Librarians should make an effort to sell their capability in RDM by visiting and talking to researchers in the comfort of their offices. Awareness campaigns about librarians' capability in RDM should be preceded by upskilling because the study established that librarians had deficits in various RDM specialities. Finally, the absence of data management polices meant that RDM activities were conducted haphazardly with no formal guidance. It is proposed that universities should work in collaboration with various research stakeholders to formulate, adopt, implement and enforce RDM policies that can clearly guide and reinforce collection, sharing, preservation and re-use of research data. These polices should also clearly stipulate mechanisms for rewarding researchers involved in data sharing.

7.5. Contribution and originality of the study

Research should add value to a particular institution, society or community by addressing pertinent issues (Wassenaar, 2006). In terms of policy, issues addressed in the study were of value to various policy makers involved in research activities including university management, librarians, government institutions, research institutions and researchers in Malawi and research grant organisations. The findings have potential to influence policy formulation for governing and reinforcing RDM initiatives in Malawi public universities based on recommendations highlighted in this chapter. In other words, the findings from this study should pave the way for further discussions in regard to the development, adoption and enforcement of RDM policies in Malawian universities. More importantly, the findings have the potential to raise awareness about the status of RDM in Malawian universities which until to date, were insufficiently researched thereby presenting various research stakeholders with best practices that are critical to the reinforcement of RDM initiatives.

In terms of theory, the study contributes valuable literature to the field of RDM especially in the context of Malawi. The evidence from the comprehensive literature search conducted in Chapter Two, points to this study as the first comprehensive research to explore RDM in Malawian universities. Beyond the borders, the study is a vital addition to the literature on RDM in Africa which until recently, was mostly available in South Africa, Kenya and sporadically in Tanzania and Uganda.

The study proposes a model which extends the value of prior models such as E-E-C Framework (Bryant et al., 2017), OAIS (Consultative Committee for Space Data Systems, 2012), the CCMF (Lyon, et al., 2011, p. 21) and the DCC Curation Lifecycle Model (Higgins, 2008). The proposed model can either be used individually or jointly with other models in implementing RDM projects or conducting digital preservation research.

7.5.1 Proposed model for research data management in Malawi

Based on the empirical research on RDM in Malawi and the literature on RDM, a contextual framework for guiding and understanding RDM in Malawian public universities is proposed. The development of the model was guided by the key findings of the study. Since the creation of new knowledge is influenced by existing knowledge, the model implicitly and explicitly adopted some aspects from the two models that guided this study. Table 34 presents the proposed elements of the model and it shows the elements that are either available or missing in the adopted models. Each element is accompanied by commentaries. Check marks (✓) and cross marks (✗) are used to show the respective availability and unavailability of a particular element/sub-element in previous models.

Table 34. Proposed elements of the model compared with adopted models

Proposed element One: Collaboration						
Sub-elements			**Check list against models adopted in this model**			
			CCMF		**DCC Curation Lifecycle Model**	
Sub-element		**Commentaries**	**✓/✗**	**Commentaries**	**✓/✗**	**Commentaries**
Researchers		The study revealed that there was lack of collaboration amongst these stakeholders in RDM.	✓	- Focusses on collaboration for intensive research production at departmental research groups, cross-research groups, national level, and international - Does not specifically focus on RDM	✓	Highlights collaboration between data creators and curators
Librarians			✗		✓	
IT personnel			✗		✗	
Directors of research in universities			✗		✗	
Government research institutions			✗		✗	
Research grant funders			✗		✗	
Proposed element Two : RDM policies						
Finances		The study revealed that absence of policies in regards to these issues affected RDM activities	✗	- Highlights legal issues that may act as barriers to sharing data and re-using - does not discuss policies on RDM	✗	This element is not implicitly or explicitly covered by the model
Promotions			✗		✗	
Data citation			✗		✗	
Data re-use			✗		✗	
Co-publication from shared data			✗		✗	
Proposed element Three: Reward system						
Promotion		The study found that lack of a proper system affected implementation of RDM in universities	✓	Through its *academic activities* it highlights that academic research community should be recognised for their research activities including RDM	✗	This element is not implicitly or explicitly covered by the model
Funds for RDM activities			✓		✗	
Data citation by re-users			✗		✗	
Proposed element Four: Infrastructure						
Software and hardware	Software	- Lack of hardware and hardware infrastructure affected RDM activities in universities - Lack of adequate funds, skilled people and time	✓	In its economic and business models, it highlights the need for investments in longitudinal data surveys and network infrastructure	✓	It highlights that data is stored in a secure manner adhering to relevant standards.
					✓	
	Hard drives		✓		✓	
	Computers		✓		✓	
	Internet		✓		✓	
	Scanners				✓	
Resources	People		✗		✗	
	Finances		✓		✗	

	Time	affected RDM activities	✗		✗	
Data repositories	University repositories		✓		✗	
	National repositories		✓		✗	
Proposed element Five: RDM competencies						
Hardware troubleshooting		The study showed that lack of skills in various RDM activities affected implementation of RDM	✗	Highlights the need for research skills in general particularly, cloud computing; visualisations; statistical analysis; simulations; data description; identification; citation	✓	The Model was developed to help the UK DCC in training creators of data, data curators, and users; to help individuals and organisations organise their digital resources; and to help organisations plan and implement the preservation of their digital assets.
Metadata			✓		✓	
Data management plans			✓		✓	
Data appraisal			✗		✓	
Data retrieval			✗		✓	
Curation lifecycle					✓	
Preservation strategies			✗		✓	
Data citation			✓		✗	
Data transformation			✗		✗	
Hardware and software installation			✗		✓	

As can be seen in Table 34, there are five components that constitute the proposed framework, namely, collaboration, RDM policies, RDM rewards, infrastructure and RDM competences. The components are aggregated to provide the final product of the model as illustrated in Figure 11. These five components should be considered when intending to examine and understand RDM activities in universities. Researchers intending to study RDM or related concepts are invited to adopt, test, critique and make further suggestions to the model.

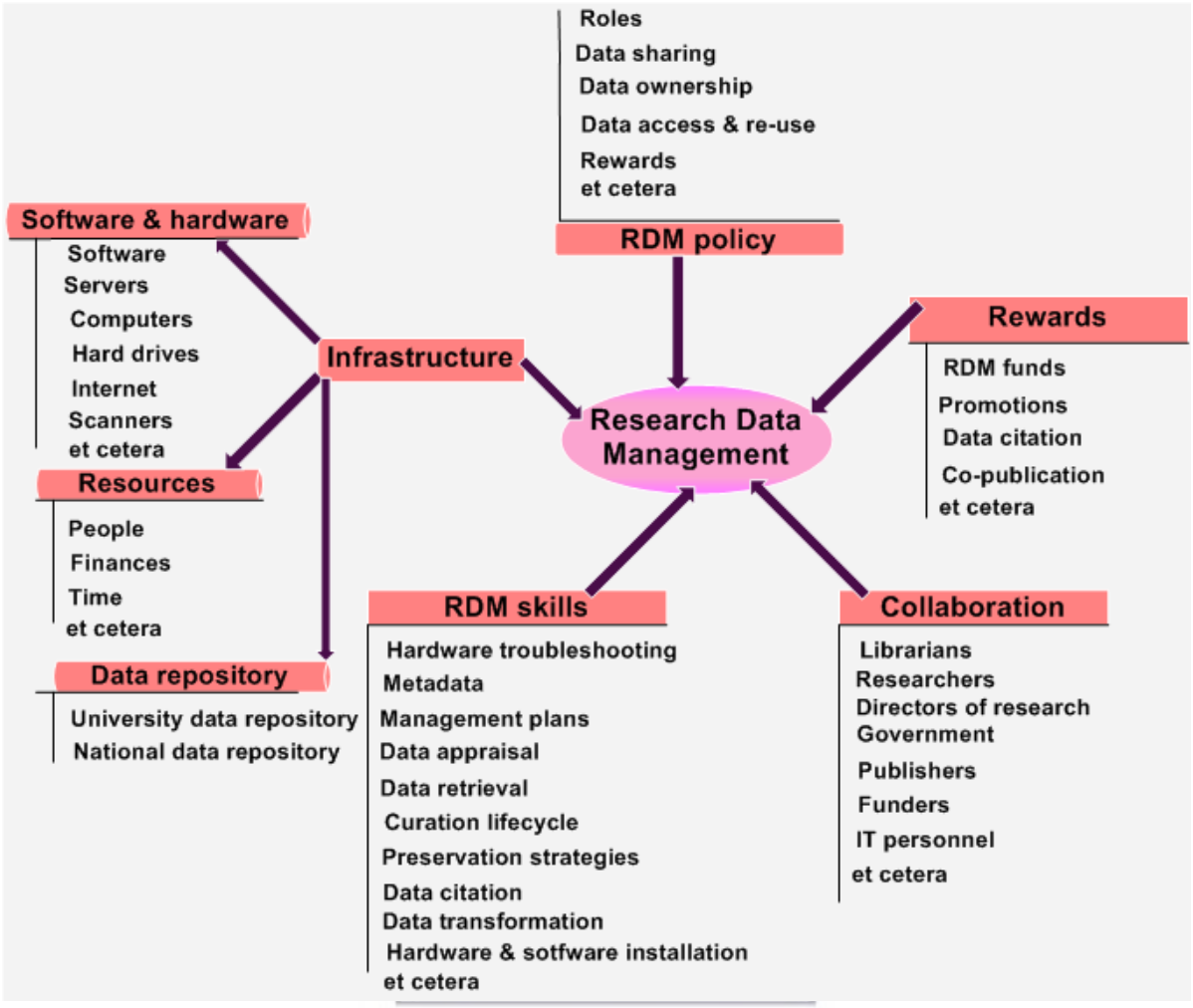


Figure 11. Proposed model for research data management

7.5.1.1 Collaboration

The aspect of collaboration has been highlighted in the CCMF (Lyon et al., 2011). However, the CCMF (Lyon et al., 2011) previewed collaboration as key in maximising research production in research intensive institutions. In this new model, it is proposed that there is a need for collaboration in RDM activities within and without the universities or research institutions. To create and adequately preserve data, the model suggests that there should be collaboration amongst various research stakeholders. In particular, Cox and Pinfield (2016) persuasively argue that libraries need to collaborate with researchers along with other key university players in pursuit of RDM activities. In more general terms, Day (2008) notes that collaboration plays a key role in the data curation process by pooling resources together. Based on the findings of the present study, collaborators could include researchers, librarians, IT personnel, directors of research in universities, government research institutions, research grant funders and other stakeholders.

Researchers

The whole data management process starts with researchers - they are the creators of data. All the four models discussed in Chapter Three: the Education-Expertise-Curation Framework (Bryant et al., 2017), OAIS (Consultative Committee for Space Data Systems, 2012), the CCMF (Lyon, et al., 2011) and the DCC Curation Lifecycle Model (Higgins, 2008) highlighted the importance of researchers in the data curation process. This model proposes that researchers have a responsibility in making sure that their data is available to the user community in line with institutional, national, journal and research grants RDM polices.

Librarians/IT officers

The debate is rife about which professionals, librarians or IT officers, are more naturally positioned to execute RDM activities. Kahn et al. (2014) observe that it appears there is little appreciation of the need for information management skills in RDM, hence the perception that IT skills are more critical. This model takes a programmatic: it considers librarians and IT staff as natural partners who invaluably and equally complement each other in RDM activities. For example, librarians are responsible for data description - assigning metadata to data using conventional standards. It is their responsibility to help researchers in preparing data management plans at the onset of the research process. IT officers bring expertise for building a networked data infrastructure that integrates software and hardware. This model proposes that librarians and IT personnel are both responsible for preparing the budgets for RDM activities. However, the model makes a cautious observation: increasingly, the role of IT personnel is being subsumed by librarians as evidenced by the increasing number of library schools incorporating IT courses in their curriculum. For instance, the role of creating web user interfaces has for decades been restricted to IT staff but librarians are equally good at performing this role.

University directors or deans of research

These are key in the formulation and enforcement of RDM policies at university level. Considering that they also represent universities on institutional, national and international research platforms, they have an opportunity to influence the formulation and enforcement of policies at these levels. Being the face of universities, they have a role to ensure that RDM policies are followed and enforced by relevant stakeholders within the university structure.

Government research institutions

These are also key in formulating and enforcing national RDM policies. The model proposes that national RDM policies have the potential to influence the effectiveness of RDM policies at university level. Since they represent governments at regional and international levels, they are best suited to lobby with diverse research stakeholders for the development and enforcement of better RDM policies that are crucial to the attainment of global free access to research data. For instance, the National Research Foundation of South Africa (Chiware & Mathe, 2016, p. 2; Koopman, 2015; Matlatse, 2016) and the National Science Foundation of USA (Cohn, 2012; Borgman, 2012) play a critical role in the popularisation of RDM activities in their respective countries.

Research grant funders and publishers

The role of publishers and research grants organisations cannot be underestimated. They hold high influence in RDM. According to Charbonneau (2013), Chen and Wu (2017), and Schumacher and VandeCreek (2015) funders increasingly require that researchers include data management plans in their research grants proposals. For instance, the EU, which is an influential international research funder, declared that beginning in 2014, all data produced from its funding should be accessible for free (European Commission, 2012). Publishers are indispensable to researchers, they are the final destination of most research activities. Their absolute necessity makes it easier for them to adopt policies that compel researchers to make their data publicly available. For instance, the International Committee of Medical Journal Editors which is an influential association of reputable publishers of medical journals requires that researchers publishing clinical trials in its member journals should have their data shared with external investigators (Ross, 2016). The model proposes that universities can directly partner with these RDM players in enforcing RDM policies. The model cautions, however, that it is easier to work with research grants organisations because they are already official natural partners of universities and research institutions. On the other hand, it is tricky working with publishers especially with the proliferation of online predatory journals; moreover, researchers are at a liberty to publish their manuscripts with publishers of their choice.

7.5.1.2 RDM policies

Polices are vital for the successful implementation of RDM activities. The model proposes the formulation of RDM policies that clearly stipulate the role of RDM research stakeholders - researchers, librarians, IT personnel, directors of research in universities, government research

institutions, research grant funders and other stakeholders. According to the present study's findings, granting researchers powers to place restrictions over data encourage them to share more; the model proposes that institutional, national and international RDM policies should clearly grant researchers reasonable control over the data they have shared. The policies should clearly provide guidance on data sharing, data re-use and reward system. For example, will the policy compel researchers to share data for all publications that the university funds in the form of research grants or partially funds by meeting publication processing fees? What about data generated in collaboration with external researchers? What about data generated with support from research grants organisations? Policies should clearly provide answers and guidance to all these questions. A bitter lesson learned, according to the present study, was that local researchers had been denied access to data they jointly generated with international collaborators - international researchers came to collaborate with them but went away with all the data and local collaborating researchers did not have an opportunity to publish from such data sets. The absence of RDM policies implies that local researchers could continue to be exploited by their international counterparts, hence the need to adopt policies that protect local universities from such malpractices. .

7.5.1.3 Reward system

RDM comes with no visible benefits to researchers who share their data. In that regard, Woolfrey (2009) argues that researchers will commit more of their time and efforts in preparing and sharing final research findings which reward them for their efforts. The present study revealed that while researchers were rewarded and recognised for publishing their findings in journals, their universities did not reward them for sharing or publishing data. As a way of inspiring researchers to partake in RDM, the model proposes that universities and research institutions should clearly stipulate how they will offer rewards to these researchers. Rewards may come in different forms but based on these findings, it is appropriate to consider issues of promotion, provision of funds for RDM activities and data citation by re-users – these should be incorporated in RDM policies. Directors or deans of research are key in implementing the rewards because they are part of university management. University RDM policies should clearly stipulate how researchers sharing their data will be rewarded. Huang et al. (2012) warn that institutional cultures, which exclude a reward system in their policies, discourage researchers from sharing data.

7.5.1.4 Infrastructure

Almost all models focusing on RDM highlight the importance of data infrastructure. Based on these findings, the model categorises RDM infrastructure into three: software and hardware; resources; and data repositories. The link between these categories of infrastructure is inseparable. The hardware and software form the component of the data storage facility (Shakeri, 2013). Two models of hardware and software preservation infrastructure can be proposed: short and long- term. Short-term data infrastructure can be centrally located and managed in the local or university library - restricted to a particular university. Long-term infrastructure can be a joint venture by universities or national research institutes to cater for the needs of geographically distributed researchers.

Software and hardware

Software is necessary for integrating the hardware, data and the users. The absence of software and hardware in the present study adversely affected RDM activities. The role of software in RDM is extensively discussed in a report by the National Science Foundation of USA (2012). The software is at the centre of data storage and management by the curators - librarians and IT personnel, and it facilitates access to the data repositories by users. The software also plays a key role in content management – it provides the interface used by the curator to upload and update data sets and it provides the user interface. The software provides a platform for assigning or editing metadata such as Dublin Core depending on the choice of the curator. The model echoes observations by the National Science Foundation of USA (2012) that development of software infrastructure is paramount for data capture and ensuring a shared and collaborative data system. The hardware in this context includes high speed servers and networks. Servers are used to store large amounts of data. To access the data remotely or on local area networks or on global networks, the data infrastructure should be well networked. Hardware may extend to user devices; universities or research institutions may purchase laptops and external hard drives for use by the curators in collecting data from researchers destined for uploading into the repository.

Resources

The model proposes resources in three categories: people, funds/money and time.

People: Human resources are crucial in the data management process. Librarians are responsible for collecting data from creators (researchers), uploading it into the repository,

managing content and providing access to the data. Librarians create and edit metadata which they assign to data as part of the data documentation process. Librarians have a role to train researchers in data management plans which are increasingly demanded by research grant organisations as part of the conditions for awarding funds. In fact, Chen and Wu (2017) sum up that librarians can provide special training related to RDM focusing on data management and sharing policies; data management plans, data discovery, retrieval and access; format, size; repository requirements; and related tools such as retrieval, recording and processing, preservation and backup for data management and sharing. Issues of data open access as advocated by the Berlin Declaration (2003) and European Commission (2012) should also be included in such training workshops. According to the findings of the present study, data sharing and re-use were affected by the failure of researchers to document the data which means it was not readable. Increasingly, librarians work in collaboration with IT staff in identifying, appraising and recommending appropriate software for running the data infrastructure.

Likewise, IT personnel play an important role in the data management process. They are commonly responsible for installing and updating software, creating ideal user interfaces and connecting the data infrastructure to the global networks. They are also responsible for all security issues regarding the data infrastructure. To perform these duties, librarians and IT personnel need various skill sets (Ng'eno, 2018) which can be obtained informally – workshops or in-house training for example, or can be obtained formally by enrolling in IT or library schools that offer RDM courses.

Funds: RDM is never inexpensive. Building a robust and dependable data infrastructure requires enough capital. Funds can come from two key sources: budgets allocation from universities or research institutions or from donors. Directors of research have a role to lobby their institutions to allocate enough funds at institutional level. Directors of research and librarians have a role to develop funding proposals for building and maintaining data infrastructure.

Time: Researchers are busy people as they are involved in other equally pressing academic activities. One of the reasons researchers did not share data in the present study was due to the lack of time. Researchers should be well rewarded for the extra time and effort they direct towards RDM activities.

Data repositories

Robust data storage repositories are fundamental in RDM. Data repositories can be divided into three forms namely, university repositories, national repositories, and international repositories. University data repositories are described by Walters and Skinner (2011) as silo-based approach to research data management or university-furnished networked (Schumacher & VandeCreek, 2015). In this approach of data management, individual universities manage and maintain control ownership over the data. On the other hand, national data repositories are described by Walters and Skinner (2011) as community-driven data sharing or geographically diverse implying that they cater for RDM needs of various research institutions in a particular country. In this context, universities may pool resources together to jointly construct a centralised repository for managing the data. Alternatively, in consultation with universities, government owned national research institutions may build a data repository which captures and manages data generated by various universities and research institutes across the country. The model proposes that national data repositories should be preferred because unlike university repositories, they lend themselves to geographical distribution and access to data is not restricted to members of any particular university or research institute. International data repositories are also increasingly becoming common. However, the concept of international data repository is beyond the scope of the proposed model; future revisions of the model may consider this aspect.

7.5.1.5 RDM competencies

The CCMF (Lyon, et al., 2011) has also discussed issues of skills and competences. Unlike the CCMF (Lyon. et al., 2011) which perceives researchers as perpetually seeking help from librarians, this model proposes that researchers need to be equipped with basic RDM skills. The literature reveals that there is a growing demand for librarians to acquire new types of skills and competencies in order to assume the new roles of digital curation (Heidorn, 2011; Newton et al., 2011; Ray, 2012). As proposed by the DCC Curation Lifecycle Model (Higgins, 2008), RDM actions include full-cycle curation activities - description and representation of information; preservation planning; curation and preservation; sequential actions - conceptualise, create and receive; appraise and select; ingest; preservation action, storing; accessing, use and reuse; and transforming. It was revealed in the present study that gaps in RDM skills amongst researchers and librarians hampered various RDM activities. Curators such as librarians need special skills for the successful implementation of the data curation

lifecycle activities. As already highlighted, curators can gain these skills informally or formally.

7.6 Suggestion for further research

The present study investigated research data management practices at two public universities in Malawi. However, there are other universities which offer specialised subjects in pure sciences and agricultural sciences which were not included. These universities include Lilongwe University of Agriculture and Natural Resources and the Malawi University of Science and Technology. Future studies should extend to these universities. Furthermore, the study focused on government funded universities leaving out privately owned universities which in their quest to fulfil their research and teaching obligations, generate research data. Hence, future studies should consider focussing on these universities and colleges.

More importantly, it has been noted that most studies on RDM in African were conducted in South Africa and Kenya implying they focused on particular countries. It could be revealing conducting a cross-country study to better understand variations influencing RDM at regional or international level.

LIST OF REFERENCES

Abbott, D. (2009). *DCC briefing papers: Introduction to curation*. Retrieved from http://www.dcc.ac.uk/resources/briefingpapers/introductioncuration/Interoperability (Accessed April 5, 2018).

Abend, G. (2008). The meaning of 'theory'. *Sociological Theory*, 26(2), 173-199.

Acord, S. K., & Harley, D. (2012). Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society*, *15*(3), 379-397. http://dx.doi.org/10.1177/1461444812465140

Adobe Systems Incorporated (2008). *Document management – Portable Document Format – Part 1: PDF 1.7.* San Jose, United States: Adobe Systems Incorporated. Retrieved from https://www.adobe.com/content/dam/acom/en/devnet/acrobat/pdfs/PDF320002008.pdf (Accessed September 17, 2017).

Aguillo, I., Bar-Ilan, J., Levene, M., & Ortega, J. (2010). Comparing university rankings. *Scientometrics*, *85*(1), 243-256. https://doi.org/110.1007/s11192-010-0190-z

Allmark, P., Boote, J., Chambers, E., Clarke, A., McDonnell, A., Thompson, A., & Tod, A. M. (2009). Ethical issues in the use of in-depth interviews: literature review and discussion. *Research Ethics*, *5*(2), 48-54. https://doi.org/10.1177/174701610900500203

Alvaro, E., Brooks, H., Ham, M., Poegel, S., & Rosencrans, S. (2011). E-science librarianship: Field undefined. *Issues in Science and Technology Librarianship*. Retrieved from https://eric.ed.gov/?id=EJ953335 (Accessed April 21, 2017).

American Library Association (ALA). (2011). Library workers: facts and figures. Retrieved from http://ala-apa.org/files/2012/03/Library-Workers-2011.pdf (Accessed January 28, 2019)

Anane-Sarpong, E., Wangmo, T., Ward, C. L., Sankoh, O., Tanner, M., & Elger, B. S. (2017).You cannot collect data using your own resources and put it on open access: Perspectives from Africa about public health data-sharing. *Developing World Bioethic*s, 1-12. http://dx.doi.org/10.10.1111/dewb.12159

Aronson, J. (1995). A Pragmatic view of thematic analysis. *The Qualitative Report, 2*(1), 1-3.

Atkins, D.E. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Retrieved from http://arizona.openrepository.com/arizona/handle/10150/106224 Accessed August 13, 2017).

Babbie, E. (2004). *The practice of social research*. (10th Ed.) Belmont, CA: Wadsworth.

Babbie, E., & Mouton, J. (2001). *The practice of social science research*. Belmont, CA: Wadsworth.

Barker, C., Pistrang, N., & Elliott, R. (2005). Foundations of quantitative measurement. *Research Methods in Clinical Psychology*, 49-71.

Beaudoin, J. E. (2011). Specters in the archive: Faculty digital image collections and the problems of invisibility. *The Journal of Academic Librarianship, 37*(6), 488–494. https://doi.org/10.1016/j.acalib.2011.07.005

Berlin Declaration (2003). Berlin Declaration on Open Access to knowledge in the sciences and humanities. Retrieved from http://oa.mpg.de/files/2010/04/Berlin-I-2.pdf (Accessed July 20, 2019).

Berry, A.J. & Otley, D.T. (2004). Case based research in accounting. In Humprey, C., & Lee, B (Eds). *The real life guide to accounting research: A behind-the-scenes view of using qualitative research methods*. Oxford: Elsevier.

Blanche, M. Durrheim, K. (1999). *Research in practice: applied methods for the social sciences*. Cape Town: University of Cape Town Press.

Blaxter, L., Hughes, C., & Tight, M. (2006). *How to research*. (3rd Ed.) Berkshire: Open University Press.

Boeije, H. (2010). *Analysis in qualitative research*. London: Sage Publications. http://dx.doi.org/10.5785/26-2-24

Bond-Lamberty, B. (2018). Data sharing and scientific impact in eddy covariance research. *Journal of Geophysical Research: Biogeosciences*, *123*(4), 1440-1443. http://dx.doi.org/10.1002/2018JG004502

Borgman, C.L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, *3*(4), 1-30. http://dx.doi.org/10.1.1.866.6966&rep=rep1&type=pdf

Borgman, C.L. (2010). Research data: Who will share what, with whom, when, and why? Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman (Accessed November 5, 2017).

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the Association for Information Science and Technology*, *63*(6), 1059-1078. http://dx.doi.org/10.1002/asi.22634

Borgman, C.L. (2015). *Big data, little data, no data: Scholarship in a networked world*. Cambridge, MA: The MIT Press.

Bouma, G., & Atkinson, B. (1995). *A handbook of social science research - a comprehensive and practical guide for students*. Oxford: Oxford University Press.

Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Watertown, MA: Pathfinder International.

Brakewood ,B., & Poldrack, R .A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage* 82, 671–676. http://dx.doi.org/10.1016/j.neuroimage.2013.02.040

Brambilla, P. (2015). *Digital Curation in the Italian context: new roles and professions for digital librarians* (Master's thesis). Oslo and Akershus University College of Applied Sciences, Italy and Norway. Retrieved from https://oda.hio.no/jspui/handle/10642/3369 (Accessed December 14, 2017).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Brewerton, P., & Millward, L. (2001). *Organisational research methods: a guide for students and researchers.* London: Sage Publications Ltd

Brown, S, Bruce, R., & Kernohan, D. (2015). Directions for research data management in UK Universities. Retrieved from http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf (Accessed October 5, 2017).

Bryant, R., Brian, L., & Malpas, C. (2017). *A tour of the research data management (RDM) service space*: *The realities of research data management, part 1*. Dublin, Ohio: OCLC Research. Retrieved from https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html (Accessed May 5, 2018).

Bryman, A. (2012). *Social research methods* (4th Ed). New York: Oxford University Press

Cahill, J. M., & Passamano, J. A. (2007). Full disclosure matters. *East Archaeol* 70, 194–196. http://dx.doi.org/10.2307/20361332

Callaghan, S. (2013). Letter from the guest content editor. *Information Standards Quarterly, 25*(3), 2-3. http://dx.doi.org/10.3789/isqv25no3.2013.01

Campbell, E.G., & Bendavid, E. (2003). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *Journal of Health Care Law Policy, 6*, 241-255.

Cantwel, P.J. (2008). Census. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp.90-93). Sage Publications.

Card, N.A. (2012). Literature Review. In N. J Salkind (Ed.), *Encyclopedia of Research Design* (pp. 726-728). Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781412961288

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy, 11*(2), 629-657. http://dx.doi.org/10.1353/pla.2011.0022

CARL Data Management Sub-Committee (2009). Research data: Unseen opportunities. Retrieved from http://carl-abrc.ca/uploads/pdfs/data_mgt_toolkit.pdf (Accessed September 4, 2017).

Cater, J.K. (2014). Using Skype for qualitative interview research. Retrieved from https://www.academia.edu/2061796/USING_SKYPE_FOR_QUALITATIVE_INTERVIEW_RESEARCH (Accessed February 6, 2018).

CHANCO (2019). *Chancellor College*. Retrieved from https://www.cc.ac.mw/ (Accessed July 25, 2019).

Chandramohan, D., Shibuya, K., Setel, P., Cairncross, S., Lopez, A. D., Murray, C. J., ... & Binka, F. (2008). Should data from demographic surveillance systems be made more widely available to researchers?. *PLoS Medicine*, 5(2), e57. http://dx.doi.org/110.1371/journal.pmed.0050057

Charbonneau, D. H. (2013). Strategies for data management engagement. *Medical reference services quarterly*, *32*(3), 365-374. https://doi.org/10.1080/02763869.2013.807089

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research, 41 (*2), 109-122. https://doi.org/10.1016/j.lisr.2019.04.004

Chawinga, W. D., & Zozie, P. A. (2016). Increasing access to higher education through open and distance learning: Empirical findings from Mzuzu University, Malawi. *The International Review of Research in Open and Distributed Learning*, *17*(4). http://dx.doi.org/10.19173/irrodl.v17i4.2409

Chawinga, W.D. & Majawa, F. (2018). An assessment of Mzuzu University Library after a fire disaster. *African Journal of Library, Archives and Information Science, 28* (2), 183-194

Chen, X., & Wu, M. (2017). Survey on the needs for chemistry research data management and sharing. *The Journal of Academic Librarianship*, *43*(4), 346-353. http://dx.doi.org/10.1016/j.acalib.2017.06.006

Chipeta, G.T., Dube, G, Chawinga, W.D., Malemia, L., & Chaura, M. G. (2019). Information-seeking behaviour of first-year undergraduate students at Mzuzu University, Malawi. *Mousaion: South African Journal of Information Studies*, *36*(1), 18. https://doi.org/10.25159/0027-2639/4931

Chiware, E., & Mathe, Z. (2016). Academic libraries' role in research data management services: a South African perspective. *South African Journal of Libraries and Information Science*, *81*(2), 1-10. http://dx.doi.org/10.7553/81-2-1563

Chivwara, N. (2013). *An analysis of the governance of higher education access in Malawi* (Master's thesis). University of the Western Cape, South Africa. Retrieved from http://etd.uwc.ac.za/handle/11394/4039 (Accessed July 24, 2019

Clement, R., Blau, A., Abbaspour, P., & Gandour-Rood, E. (2017). Team-based data management instruction at small liberal arts colleges. *IFLA Journal*, *43*(1), 105-118. http://dx.doi.org/10.1177/0340035216678239

Connaway, L.S., & Powell, R.R. (2004). *Basic research methods for librarians* (10th Ed). Greenwich, Conn: Libraries Unlimited.

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023-4038. http://dx.doi.org/10.1098/rsta.2010.0165

Cohn, J. P. (2012). DataONE opens doors to scientists across disciplines. *BioScience, 62*(11), 1004-1004. http://dx.doi.org/10.1525/bio.2012.62.11.16

Constantopoulos, P., Dallas, C., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., & Papatheodorou, C. (2009). DCC&U: An extended digital curation lifecycle model. *International Journal of Digital Curation, 4*(1), 34-45. http://dx.doi.org/10.2218/ijdc.v4i1.76

Consultative Committee for Space Data Systems (CCSDS). (2002). Reference model for an open archival information system (OAIS) (recommended practice No. CCSDS 650.0-M-2). Retrieved from http://public.ccsds.org/publications/archive/650x0m2.pdf (Accessed April 8, 2018)

Cooper, M. (2007). Sharing data and results in ethnographic research: Why this should not be an ethical imperative. *Journal of Empirical Research on Human Research Ethics*, *2*(1), 3-19. http://dx.doi.org/10.1525/jer.2007.2.1.3

Corrall, S. (2012). Roles and responsibilities: Libraries, librarians and data. In G. Pryor (Ed.). *Managing Research Data (*pp.105-133). London: Facet.

Corvellec, H. (2013). *What is theory? Answers from the social and cultural sciences*. Copenhagen: Copenhagen Business School Press DK.

Costello, M.J. (2009). Motivating online Publication of data. *BioScience*, 59, 418–427. http://dx.doi.org/10.1525/bio.2009.59.5.9

Cox, A.M., & Pinfield S (2016) Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science 46*(4), 299-316. http://dx.doi.org/10.1177/0961000613492542

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-131

Creswell, J.W. (2003). *Research design: qualitative, quantitative and mixed methods approaches*. Los Angeles: Sage Publication.

Creswell, J.W. (2009). *Research design: qualitative, quantitative and mixed methods approach*. Los Angeles: Sage Publications.

Creswell, J.W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches*. Los Angeles: Sage.

Creswell, J.W., & Creswell, J.D. (2018). *Research design: qualitative, quantitative and mixed methods approaches*. London: Sage Publications, Inc.

Creswell, J.W., & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Los Angeles: Sage publications.

Creswell, J.W., & Plano Clark, V.L. (2011). *Designing and conducting mixed methods research*. Los Angeles: Sage publications.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research proces*s. London: Sage Publications.

Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PloS One, 12*(12), e0189288. https://doi.org/10.1371/journal.pone.0189288

Dai, S. Q., Li, H., Xiong, J., Ma, J., Guo, H. Q., Xiao, X., & Zhao, B. (2018). Assessing the extent and impact of online data sharing in eddy covariance flux research. *Journal of Geophysical Research: Biogeosciences, 123*(1), 129-137. http://dx.doi.org/10.1002/2017JG004277

Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard Business Review, 90*(5), 70-76.

Davidson, J., Jones, S., Molloy, L., & Kejser, U.B. (2014). Emerging good practice in managing research data and research information in UK Universities. *Procedia Computer Science, 33*, 215-222. https://doi.org/10.1016/j.procs.2014.06.035

Day, M. (2008). Toward distributed infrastructures for digital preservation: the roles of collaboration and trust. *International Journal of Digital Curation, 3*(1), 15-28. http://dx.doi.org/10.2218/ijdc.v3i1.39

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research, 12*(2), 279-289.

Davies, P. (2006). Informed consent. In V. Jupp (Ed.). *The Sage dictionary of social research* (pp. 310-322). London: Sage.

De Vos, A.S., Strydom, H., Fouche, C.B., & Delport, C.S.L. (2005). *Research at grass roots: for the social sciences and human service professions* (3rd Ed). Pretoria: Van Schaik Publishers.

Decrop, A. (1999). Triangulation in qualitative tourism research. *Tourism Management, 20*(1),157-161.

Denny, S. G., Silaigwana, B., Wassenaar, D., Bull, S., & Parker, M. (2015). Developing ethical practices for public health research data sharing in South Africa: The views and experiences from a diverse sample of research stakeholders. *Journal of Empirical Research on Human Research Ethics, 10*(3), 290-301. http://dx.doi.org/10.1177/1556264615592386

Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd Ed). Maidenhead: Open University Press.

Denzin, N. K., & Lincoln, Y. S. (2011). Introduction: The discipline and practice of qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.). *The SAGE Handbook of Qualitative Research* (pp. 1-19). Thousand Oaks: Sage.

Digital Curation Centre (2008). What are the benefits of this model? Retrieved from http://www.dcc.ac.uk/node/9561(Accessed May 15, 2018).

Digital Curation Centre (2011). *Data management in perspective: The career profile of data managers.* Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/webfmsend/526 (Accessed April 10, 2016).

Digital Curation Centre (2017).What is digital curation? Retrieved from http://www.dcc.ac.uk/digital-curation/what-digital-curation (Accessed December 12, 2017).

Divala, J.J. (2013). Rethinking higher education access in Malawi: dilemmas in achieving a just system. Retrieved from http://www.ndr.mw:8080/xmlui/bitstream/handle/123456789/512/Rethinking%20higher%20education%20access%20in%20Malawi%20dilemmas%20in%20achieving%20a%20just%20system..pdf?sequence=1 (Accessed May 23, 2018).

Doorn, P., Dillo, I. & Van Horik, R. (2013). Lies, damned lies and research data: can data sharing prevent data fraud? *International Journal of Digital Curation, 8*, 229-243. http://dx.doi.org/10.2218/ijdc.v8i1.256

Dryad (2010). Dryad sustainability plan. Retrieved from https://nescent.org/wg/dryad/images/b/bf/Beagrie_suppdata_report_apr10.pdf (Accessed October 10, 2017).

Dublin Core Metadata Initiative. 2004. Dublin core metadata element set, version 1.1: Reference description. Retrieved from http://dublincore.org/documents/dces/ (Accessed October 19, 2017).

Dunckley, M., Reshef, D., Conway, E., & Giaretta, D. (2010). Using XFDU for CASPAR information packaging. *OCLC Systems and Services*, *26*(2), 80-93. https://doi.org/10.1108/10650751011048452

Durrheim, K., & Painter, D. (2006). Collecting qualitative data: sampling and measuring. In: Terre Blanche, M., Durrheim, K. and Painter, D., (Eds), *Research in practice: applied methods for the social sciences*. University of Cape Town: Cape Town.

Edmonds, W A., & Kennedy, T.D. (2013). *An applied reference guide to research designs: Quantitative, qualitative and mixed methods.* Sage Publications.

Egger, A. (2006). Shortcomings of the reference model for an Open Archival Information System (OAIS). *IEEE TCDL Bulletin, 2(2).* Retrieved from http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html (Accessed on May 11, 2018).

Elsayed, A. M., & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA Journal*, 1-9. http://dx.doi.org/10.1177/0340035218785196

Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (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. https://doi.org/10.1016/j.ecoinf.2012.03.004

European Commission (2012). Scientific data: open access to research results will boost Europe's innovation capacity. Retrieved from http://europa.eu/rapid/press-release_MEMO-12-565_en.htm?locale=en (Accessed November 23, 2017).

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PloS one,* 10(2): e0118053. https://doi.org/10.1371/journal.pone.0118053

Fetters, M.D., Curry, L.A., & Creswell, J.W. (2013). Achieving integration in mixed methods designs-principles and practices. *Health Services Research, 48,* 2134-2156. http://dx.doi.org/10.1111/1475-6773.12117

Fienberg, S.E. Martin, M .E., & Straf, M .L. (1985). *Sharing Research Data*. Washington, D.C.: National Academy Press.

Fry, J., Lockers, S., Oppenheim, C., Houghton, J., & Rasmussen, B. (2008). *Identifying benefits arising from the curation and open sharing of research data produced by UK higher education and research institutes*. UK: Loughborough University/Centre for Centre for Strategic Economic Studies.

Gabridge, T. (2009). The last mile: Liaison roles in curating science and engineering research data. *Research Library Issues,* 265, 15-21. https://doi.org/10.29242/rli.265.4

German Network for Bioinformatics Infrastructure (2017*). de.NBI Systems Biology Service Center (de.NBI-SysBio).* Retrieved from https://www.denbi.de/denbi-sysbio (Accessed September 1, 2017).

Given, L.M. (2008). *The Sage encyclopaedia of qualitative research method.* Newbury Park, California: Sage Publications.

Glaeser, P. S. (1990). *Scientific and technical data in a new era*. New York, NY: Hemisphere Publishing Corporation.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607.

Gordon-Murnane, L. (2012). Big data: A big opportunity for librarians. *Online, 36*(5), 30-34

Grant, C., & Osanloo, A. (2014). Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house". *Administrative Issues Journal, 4*(2), 12-26. http://dx.doi.org/10.5929/2014.4.2.9

Gray, D.E. (2009). *Doing social research in the real world*. Los Angeles: Sage.

Greene, J.C. Kreider, H., & Mayer, E. (2005). Combining qualitative and quantitative methods in social inquiry. In Somekh, B., & Lewin, C. (Ed.). *Research methods in the social sciences*. Sage: London.

Greenstein, D., & Smith, A. (2003). Digital preservation in the United States: Survey of current research, practice, and common understandings. In A, Smith (Ed.), *New-model scholarship: How will it survive?* (pp. 40-48). Washington: Council on Library and Information Resources.

Guedon, J.C. (2015). Open data and science: Towards optimizing the research process. Retrieved from https://www.dataone.org/webinars/open-data-and-science-towards-optimizing-research-process (Accessed November 4, 2017).

Habert, B., & Huc, C. (2010). Building together digital archives for research in social sciences and humanities. *Social Science Information*, *49*(3), 415-443. http://dx.doi.org/10.1177/0539018410371570

Hair, J.F., Bush, R.P., & Ortinau, D.J. (2003). *Marketing research within a changing information environment*. McGraw-Hill Irwin: Boston.

Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., & Donatuto, J. (2013). Conducting research with tribal communities: sovereignty, ethics, and data-sharing issues. *Environmental Health Perspectives, 120*(1), 6-10. http://dx.doi.org/10.1289/ehp.1103904

Harvey, R. (2010). *Digital curation: A how-to-do-it manual*. London: Facet.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research*, *3*(3), 118-126.

Heidorn, P.B. (2011). The emerging role of libraries in data curation and E-science. *Journal of Library Administration, 51*, 662-672. http://dx.doi.org/10.1080/01930826.2011.601269

Henn, M., Weinstein, M., & Foard, N. (2009). *A critical introduction to social research*. Sage Publications.

Hennink, M.M., Hutter, I., & Bailey, A. (2011). *Qualitative research methods.* London: Sage Publications.

Henning, E. (2004). *Finding your way in qualitative research*. Pretoria: van Schalk Publishers.

Henty, M. (2008). Dreaming of data: The library's role in supporting e-research and data management. In *Australian Library and Information Association biennial conference*, Alice Springs, Australia, 2nd - 5th September 2008. Retrieved from http://apsr.anu.edu.au/presentations/henty_alia_08.pdf (Accessed January 14, 2017).

Hey, T., & Trefethen, A.E. (2003). The data deluge: *An eScience perspective*. Retrieved from https://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf (Accessed December 12, 2017).

Hey, T., & Hey, J. (2006). E-Science and its implications for the library community. *Library Hi Tech*, *21*(4), 515-528. http://dx.doi.org/10.1108/07378830610715383

Heyvaert, M., Maes, B., & Onghena, P. (2013). Mixed methods research synthesis: definition, framework, and potential. *Quality & Quantity, 47*(2), 659-676. http://dx.doi.org/10.1007/s11135-011-9538-6

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, *3*(1), 134-140. http://dx.doi.org/10.2218/ijdc.v3i1.48

Higgins, S. (2011). Digital curation: the emergence of a new discipline. *The International Journal of Digital Curation*, *2*(6), 78-88. http://dx.doi.org/10.2218/ijdc.v6i2.191

Housewright, R., Schonfeld, R.C., & Wulfson, K. (2013). *Ithaka S+R | JISC | RLUK UK Survey of Academics 2012.* Research Libraries: New York.

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70-85. https://doi.org/10.1177/25152459177518

Huang, X., Hawkins, B.A., Lei, F., Miller, G. L., Favret, C., Zhang, R., & Qiao, G. (2012). Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conservation Letters, 5*(5), 399-406. http://dx.doi.org/10.1111/j.1755-263X.2012.00259.x

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., & Cuellar, A. A. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics, 19*(4), 524-531. https://doi.org/10.1093/bioinformatics/btg015

Hyams, E. (2008). Data librarianship: A gap in the market. Retrieved from http://www.era.lib.ed.ac.uk/handle/1842/2499 (Accessed November 20, 2017).

Israel, G.D. (2013). *Determining sample size*. University of Florida: Institute of Food and Agricultural Sciences.

Jaccard, J., & Jacoby, J. (2010). *Theory construction and model-building skills: A practical guide for social scientists*. New York: Guilford Press.

Jao, I., Kombe, F., Mwalukore, S., Bull, S., Parker, M., Kamuya, D., ... & Marsh, V. (2015). Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations. *PloS one*, *10*(9), e0135545. http://dx.doi.org/10.1371/journal.pone.0135545

Jeng, W., He, D., & Chi, Y. (2017). Social science data repositories in data deluge: A case study of ICPSR's workflow and practices. *The Electronic Library*, *35*(4), 626-649. https://doi.org/10.1108/EL-11-2016-0243

Johnson, R.B., & Onwuegbuzie, A.J. (2004) Mixed methods research: a research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26. http://dx.doi.org/10.3102/0013189X033007014

Joint Information Systems Committee (2016). Research at risk. Retrieved from https://www.jisc.ac.uk/rd/projects/research-at-risk (October 10, 2017).

Jonker, J., & Pennink, B. (2010). *The essence of research methodology: A concise guide for master and PhD students in management science*. Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-540-71659-4

Kahn, M, Higgs, R, Davidson, J., & Jones, S. (2014) Research data management in South Africa: How we shape up. *Australian Academic & Research Libraries*, *45*(4), 296-308. http://dx.doi.org/10.1080/00048623.2014.951910

Kalof, L., Dan, A., & Dietz, T. (2008). Essentials of social research. New York: McGraw-Hill.

Kaluwa, B. (2010). Malawi industrial performance: What else is there? *In Malawi Knowledge Network,* 10th March, 2010. Sun Hotel, Malawi.

KCN (2019). Kamuzu College of Nursing. Retrieved from http://www.kcn.unima.mw/ (Accessed July 20, 2019).

Kaye, J., Terry, S. F., Juengst, E., Coy, S., Harris, J. R., Chalmers, D., ... & Bezuidenhout, L. (2018). Including all voices in international data-sharing governance. *Human Genomics*, *12*(1), 13. https://doi.org/10.1186/s40246-018-0143-9

Kim, J., Warga, E., & Moen, W. (2013). Competencies required for digital curation: An analysis of job advertisements. *International Journal of Digital Curation, 8*(1), 66-83. http://dx.doi.org/10.2218/ijdc.v8i1.242

Kirigia, J. M., Kathyola, D. D., Muula, A. S., & Ota, M. M. O. (2015). National health research system in Malawi: dead, moribund, tepid or flourishing? *BMC Health Services Research, 15*(1), 1-14. https://doi.org/10.1186/s12913-015-0796-1

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3-14. http://dx.doi.org/10.1177/0961000615616450

Koopman, M. M. (2015). *Data archiving, management initiatives and expertise in the Biological Sciences Department* (Master's thesis). University of Cape Town, Cape Town, South Africa. Retrieved from http://open.uct.ac.za/handle/11427/13656 (Accessed June 20, 2017).

Kowalczyk, S, & Shankar, K. (2011). Data sharing in the sciences. *Annu Rev Inf Sci Technol* 45, 247-294. http://dx.doi.org/10.1002/aris.2011.1440450113

Kumar, S. (2010). Undergraduate perceptions of the usefulness of Web 2.0 in higher education: survey development. Retrieved from http://plaza.ufl.edu/swapnak/ecel09Kumar.pdf (Accessed March 23, 2017).

Kunze, J., Cruse, P., Hu, R., Abrams, S., Hastings, K., Mitchell, C., & Schiff, L.R. (2011). eScholarship: Practices, trends, and recommendations in technical appendix usage for selected data-intensive disciplines. Retrieved from http://escholarship.org/uc/item/9jw4964t (Accessed December 10, 2017)

Krier, L., & Strasser, C.A. (2014). *Data management for libraries: A LITA guide.* Chicago: American Library Association.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Beverly Hills, California: Sage Publications.

Lapan, S. D., Quartaroli, M. T., & Riemer, F.J. (2012*). Qualitative research: an introduction to methods and designs.* San Francisco: Jossey-Bass.

Latham, B. (2017). Research data management: Defining roles, prioritizing services, and enumerating challenges. *The Journal of Academic Librarianship*, *43*(3), 263-265. https://doi.org/10.1016/j.acalib.2017.04.004

Laughlin, R. (1995). Empirical research in accounting: alternative approaches and a case for "middle-range" thinking. *Accounting, Auditing and Accountability Journal*, *8*(1), 63-87

Laughton, P. (2012). OAIS functional model conformance test: a proposed measurement. *Program*, *46*(3), 308-320. https://doi.org/10.1108/00330331211244850

Lavoie, B.F. (2004). The open archival information system reference model: introductory guide. *Microform & Imaging Review*, *33*(2), 68-81. https://doi.org/10.1515/MFIR.2004.68

Lee, C.A. (2010). Open archival information system (OAIS) reference model. In: *Encyclopedia of Library and Information Sciences* (pp. 4020-4030). Oxford: Taylor & Francis. https://doi.org/10.1081/E-ELIS3-120044377

Lee, D.J. (2015). *Research data curation practices in institutional repositories and data identifiers* (Unpublished doctoral dissertation). Florida State University, USA.

Leech, N.L., & Onwuegbuzie, A.J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265-275. http://dx.doi.org/10.1007/s11135-007-9105-3

Leedy, P.D., & Ormrod J.E. (2005). *Practical research: planning and designing*. Pearson: Upper Saddle River.

Lesser, E. L. (2000). Leveraging social capital in organisations. In E.L. Lesser (Ed.). *Knowledge and social capital: foundations and applications* (pp.3-16). Woburn, MA: Butterworth-Heinemann. https://doi.org/10.1016/B978-0-7506-7222-1.50004-0

Lewis, M. (2010). Libraries and the management of research data. In S. McKnight (Ed.), *Envisioning future academic library services initiatives, ideas and challenges* (pp. 145-168). London, England: Facet Publishing.

Library and Information Studies Centre (2017). *UCT first in Africa to offer MPhil specialising in Digital Curation*. http://www.lisc.uct.ac.za/digitalcuration_mphil (Accessed December 4, 2017).

Library of Congress. (2013). MARC Standards. Retrieved from http://www.loc.gov/marc/ (Accessed September 5, 2017).

Lombe, F.B. (2013). *Higher education and national development: the response of higher education institutions in Malawi (*Doctoral dissertation). University of the Western Cape: South Africa. Retrieved from http://nrfnexus.nrf.ac.za/handle/20.500.11892/102234 (Accessed May 23, 2018).

LUANAR (2019). LUANAR graduate tracer study for cohorts 2013 and 2014. Lilongwe: LUANA.

Luck, L., Jackson, D., & Usher, K. (2006). Case study: a bridge across the paradigms. *Nursing inquiry*, *13*(2), 103-109. http://dx.doi.org/10.1111/j.1440-1800.2006.00309.x

Lunenburg, F. C., & Irby, B. J. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences*. Corwin press. (Accessed February 18, 2017).

Lyon, L. (2012). The informatics transform: Re-engineering libraries of the data decade. *International Journal of Digital Curation, 7*(1), 126-138. https://doi.org/10.2218/ijdc.v7i1.220

Lyon, L., Ball, A., Duke, M., & Day, M. (2011). Community Capability Model Framework. Retrieved from http://www.academia.edu/download/30836993/CCMDIRWhitePaper24042012.pdf

Lyon, L., Patel, M., & Takeda, K. (2014). Assessing requirements for research data management support in academic libraries: Introducing a new multi-faceted capability tool. Retrieved from http://ozk.unizd.hr/proceedings/index.php/lida/article/view/121 (Accessed July 4, 2018).

The Malawi Growth and Development Strategy (2017). Building a Productive, Competitive and Resilient Nation. Retrieved from https://cepa.rmportal.net/Library/government-publications/the-malawi-growth-and-development-strategy-mgds-iii/view (Accessed December 4, 2017).

Marinai, S. (2009). Metadata extraction from PDF papers for digital library ingest. In *Proceedings of the 10th International Conference on Document Analysis and Recognition* (pp. 251- 255), 26th-29th July, 2009. Barcelona, Spain: IEEE. https://doi.org/10.1109/ICDAR.2009.232

Malawi Growth Development Strategy (2017). Building a productive, competitive and resilient nation. Retrieved from https://www.afidep.org/download/policy_guidelines/Malawi-Growth-_-Development-Strategy-MGDS-III-2017-2022-Final-Version-copy.pdf (Accessed March 3, 2019).

Matlatse, R. L. (2016). *An evaluation of a structured training event aimed at enhancing the research data management (RDM) knowledge and skills of library and information science (LIS) professionals in South African higher education institutions (HEIs)* (Unpublished master's thesis). University of Pretoria, South Africa.

Matveev, A.V. (2002). The advantages of employing quantitative and qualitative methods in intercultural research: practical implications from the study of the perceptions of intercultural communication competence by American and Russian managers. *Theory of Communication and Applied Communication, 1*(6), 59-67.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (3rd Ed). Sage London: Publications.

Mennes, M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2013). Making data sharing work: The FCP/INDI experience. *NeuroImage, 82*, 683-691 https://doi.org/10.1016/j.neuroimage.2012.10.064

McDonough, J.P. (2012). Knee-deep in the data: practical problems in applying the OAIS Reference Model to the preservation of computer games. In: *System Science (HICSS), 45th Hawaii International Conference* (pp. 1625-1634). 4th-7th January, 2012. Maui, Hawaii USA: IEEE. https://doi.org/10.1109/HICSS.2012.1

McMillan, J.H. (2004). *Educational research: Fundamentals for the consumer* (4th Ed.) New York: Pearson.

McNamara, C. (1999). General guidelines for conducting research interviews. Retrieved from http://managementhelp.org/businessresearch/interviews.htm (Accessed December 19, 2017).

Microsoft (2006). How to minimize metadata in Word 2003. Retrieved from http://support.microsoft.com/kb/825576 (Accessed December 4, 2017).

Ministry of Health (2012). National health research agenda. Retrieved from https://www.ncst.mw/wp-content/uploads/2016/01/National-Health-Research-Agenda-NHRA.pdf (Accessed January 7, 2018).

Mitchell, R., Rose, P., & Asare, S. (2018). *Research in African universities to inform the Sustainable Development Goal for Education: Visibility, gaps and future priorities.* Centre for Education and International Development (CEID) Annual Conference, Institute of Education: London.

Monastersky, R. (2013). Publishing frontiers: The library reboot. *Nature*, *495* (7442), 430-432.

Morgan, D.L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*(1), 48-76. http://dx.doi.org/10.1177/2345678906292462.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Mouton, J. (2001). *How to succeed in your master's and doctoral studies: a South African guide and resource book.* Pretoria: Van Schaik.

Muijs, D. (2010). *Doing quantitative research in education with SPSS*. London: Sage. http://dx.doi.org/10.4135/9781849203241

Mundel, T. (2014). Knowledge is power: Sharing information can accelerate global health impact. Retrieved from https://www.impatientoptimists.org/Posts/2014/11/Knowledge-is Power#.WizthFWWbIU (Accessed December 10, 2017).

Musgrave, S. (2003). The metadata dynamic: ensuring data has a long and healthy life. In

R., Currall, J., Francis, J., Gerrard, L., Khan, R., Macer, T., Rigg, … A. Westlake (Eds.), *The impact of new technology on the survey process* (pp.110-119). Retrieved from http://www.asc.org.uk/publications/proceedings/ASC2003Proceedings.pdf (Accessed August 26, 2017).

MUST (2019). MUST at a glance. https://www.must.ac.mw/about/ (Accessed July 26, 2019).

Mzuni (2017). *Mzuni annual report*. Mzuni: Malawi.

Nalumaga, R. (2017). iSchools and Africa: Trends and developments. *Bulletin of the Association for Information Science and Technology, 42*(4),17-21. https://doi.org/10.1002/bul2.2016.1720420406

National Commission for Science and Technology (2013a). The national research agenda in the social sciences and humanities. Retrieved from https://www.ncst.mw/wp-content/uploads/2016/01/NRAiSSH.pdf (Accessed December 10, 2017).

National Commission for Science and Technology (2013). The national research agenda in the social sciences and humanities. Retrieved from http://www.nche.ac.mw/index.php/component/k2/item/38-call-for-applications-2017-18-undergraduatie-intake-into-public-universities (Accessed February 5, 2018).

National Council for Higher Education (2018). *Accreditation status of institutions*. Available online at: https://www.nche.ac.mw/index.php/accreditation/institutions-accreditation-status (Accessed March 20, 2018).

National Council for Higher Education (2018a). *Accredited higher education institutions (as of April, 2018).* NCHE: Malawi.

National Science Board. (2005). Long-lived digital data collections: Enabling research and education in the 21st Century. Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/ (Accessed August 15, 2017).

National Science Foundation of USA (2012). *Sustainable digital data preservation and access network partners (DataNet).* Arlington, VA: NSF. Retrieved from https://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm (Accessed December 3, 2017).

National Statistical Office (2008). *Population and housing census.* Zomba, Malawi: National Statistical Office.

National Statistical Office and ICF (2017). *Malawi Demographic and Health Survey 2015-16*.Zomba, Malawi, and Rockville, Maryland, USA.NSO and ICF.

Nelson, B. (2009). *Data sharing: Empty archives*. *Nature, 461,* 160-163. https://doi.org/10.1038/461160a

Neuman, WL. (2000). *Social research methods: qualitative and quantitative approaches*. Boston, MA: Allyn and Bacon. New York: Routledge.

Neuman, W.L. (2011). *Social research methods: Qualitative and quantitative approaches*. Boston: Pearson/Allyn and Bacon. http://dx.doi.org/10.2307/3211488

Newton, M.P., Miller, C.C., & Bracke, M.S. (2011). Librarian roles in institutional repository data set collecting: Outcomes of a research library task force. *Collection Management, 36*(1), 53-67. https://doi.org/10.1080/01462679.2011.530546

Ng'eno, E.J. (2018). *Research data management in Kenya's agricultural research institutes* (Unpublished D. Phil. Thesis). University of KwaZulu-Natal KwaZulu-Natal, South Africa.

Noonan, D., & Chute, T. (2014). Data curation and the university archives. *American Archivist, 77*(1), 201-240. https://doi.org/10.17723/aarc.77.1.m49r46526847g587

Noko, P., & Ngulube, P. (2015). A vital feedback loop in educating and training archival professionals a tracer study of records and archives management graduates in Zimbabwe. *Information Development, 31*(3), 270-283. https://doi.org/10.1177/0266666913510308

Ogburn, J.L. (2010). The imperative for data curation. *Portal: Libraries & the Academy, 10*(2), 241-246. https://doi.org/10.1353/pla.0.0100

Ohaji, I. K. (2016). *Research Data Management: An exploration of the data librarian role in New Zealand research organisations* (Doctoral thesis). Victoria University of Wellington, Australia. Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/5657 (Accessed July 21, 2017).

Oliver, D.G., Serovich, J.M., & Mason, T.L. (2005). Constraints and opportunities with interview transcription: Toward reflection in qualitative research. *Social Forces, 48*(2), 1273-1298.

Onwuegbuzie, A.J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12*(2), 281-316.

Onwuegbuzie, A.J., & Leech, N.L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and quantitative research methodologies. *International Journal of Social Research Methodology, 8*(5), 375-387. http://dx.doi.org/10.1080/13645570500402447

Onwuegbuzie, A.J., & Leech, N.L. (2007). A call for qualitative power analyses. *Quality & Quantity, 41*(1), 105-121. http://dx.doi.org/10.1007/s11135-005-1098-1

Ormston, R., Spencer, L., Barnard, M., & Snape, D. (2014). The foundations of qualitative research. In J. Ritchie, J. Lewis, C. Nicholls, & R. Ormston (Eds.). *Qualitative research* practice. *A guide for social science students and researchers* (pp. 1-25). London: Sage Publications.

Parker, L.D. (2003). Qualitative research in accounting and management: The emerging agenda. *Journal of Accounting and Finance* (2), 15-30

Parr, C.S. (2007). Open sourcing ecological data. *BioScience*, 57, 309. https://doi.org/10.1641/B570402

Patterton, L., Bothma, T. J., & van Deventer, M. J. (2018). From planning to practice: an action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science, 84*(2), 14-26. https://doi.org/10.7553/84-2-1761

Paton, N.W. (2008). Managing and sharing experimental data: standards, tools and pitfalls. *Biochem Soc Trans, 36,* 33-36. https://doi.org/10.1042/BST0360033

Patton, M.Q. (1990). *Qualitative research and evaluation methods*. Newbury Park, CA: Sage.

Patton, M.Q. (2002). *Qualitative research and evaluation methods*. Sage: Thousand Oaks.

Permanent Access to the Records in Europe (2009). *First insights into digital preservation of research output in Europe*. http://libereurope.eu/wp-content/uploads/PARSE-Insight_D3-5InterimInsightReport_final.pdf (Accessed December 9, 2017).

Pickard, A.J. (2007). *Research methods in information*. London: Facet Publishing.

Pisani, E., & AbouZahr, C. (2010). Sharing health data: good intentions are not enough. *Bulletin of the World Health Organization*, *88*(6), 462-466. http://dx.doi.org/10.2471/BLT.09.074393

Pitt, M. A., & Tang, Y. (2013). What should be the data sharing policy of cognitive science? *Topics in Cognitive Science*, *5*(1), 214-221. http://dx.doi.org/10.1111/tops.12006

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. PloS one, 6(7), e18657. https://doi.org/10.1371/journal.pone.0018657

Pogue, D. (2009). Should you worry about data rot? Retrieved from http://www.nytimes.com/2009/03/26/technology/personaltech/26pogue-%20email.html?_r=2& (Accessed September, 2017).

The Polytechnic (2019). Academic sections. Retrieved from https://www.poly.ac.mw/ (Accessed July, 2019).

Prado, J. C., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, *63*(2), 123-134. https://doi.org/10.1515/libri-2013-0010

Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.). *Managing research data* (pp. 1-16). London: Facet.

Pryor. G, & Donnelly, M. (2009). Skilling up to do data: Whose role, whose responsibility, whose career? *International Journal of Digital Curation, 4*(2), 158-170. https://doi.org/10.2218/ijdc.v4i2.105

Qin, J., Crowston, K., & Lyon, L. (2016). Building capabilities for sustainable research data. Management Practices. Retrieved from https://www.asist.org/files/meetings/am16/Building_Capabilities.pdf (Accessed December 10, 2017).

Raju, J. (2013). The LIS school in the ICT age: a casualty, or catalyst for making a" cross-border" shift of a different type? - the case of South Africa. Retrieved from http://library.ifla.org/177/ (Accessed April 2, 2019).

Ravitch, S.M., & Riggan, M. (2016). *Reason & rigor: How conceptual frameworks guide research*. London: Sage Publications.

Ray, J. (2012). The rise of digital curation and cyberinfrastucture. *Library Hi Tech, 30*(4), 604-622. http://dx.doi.org/10.1108/07378831211285086

Research Information Network. (2008). Stewardship of digital research data: a framework of principles and guidelines. Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines (Accessed December 5, 2017).

Rowhani-Farid, A., & Barnett, A.G. (2016). Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ open*, *6*(10), e011784. http://dx.doi.org/10.1136/bmjopen-2016-011784

Ross, J. S. (2016). Clinical research data sharing: What an open science world means for researchers involved in evidence synthesis. *Systematic Reviews, 5*(1), 159. http://dx.doi.org/10.1186/s13643-016-0334-1

Royal Society. (2012). *Science as an open enterprise: open data for open science*. London: The Royal Society Science Policy Centre.

Rubin, H.J., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data.* Thousand Oaks, California: Sage Publications. http://dx.doi.org/10.4135/9781452226651

Rusbridge, C., Burnhill, P., Ross, S., Buneman, P., Giaretta, D., Lyon, L., & Atkinson, M. (2005). The digital curation centre: a vision for digital curation. In *Local to global data tnteroperability-challenges and technologies, 2005* (pp. 31-41). Sardinia, Italy: IEEE. Retrieved from http://eprints.gla.ac.uk/33612/1/33612.pdf (Accessed October 14, 2017).

Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organisations*. New York: Basic Books

Sandberg, J. (2005). How do we justify knowledge produced within interpretive approaches? *Organization Research Methods*, *8*(1), 41-68. http://dx.doi.org/10.1177/1094428104272000

Sansone, S. A., & Rocca-Serra, P. (2012). On the evolving portfolio of community-standards and data sharing policies: turning challenges into new opportunities. *GigaScience, 1*(1), 10. http://dx.doi.org/10.1186/2047-217X-1-10

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. London: Pearson Education.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students.* (6th Ed.) London: Prentice Hall.

Savage, C. J, & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE* (4), e7078. http://dx.doi.org/10.1371/journal.pone.0007078

Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open data in global environmental research: The Belmont Forum's open data survey. *PloS One*, *11*(1), e0146695. http://dx.doi.org/10.1371/journal.pone.0146695

Schöpfel, J., Ferrant, C., André, F., & Fabre, R. (2018). Research data management in the French National Research Center (CNRS). *Data Technologies and Applications*, *52*(2), 248-265. http://dx.doi.org/110.1108/DTA-01-2017-0005

Schumacher, J., & VandeCreek, D. (2015). Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation*, *10*(2), 96-109. http://dx.doi.org/10.2218/ijdc.v10i2.32

Schutt, R.K. (2006). *Investigating the social world: the process and practice of research* (5th Ed.). Thousand Oaks, California: Sage Publications.

Scott, M. (2014). *Research data management (*Doctoral dissertation). University of Southampton, United Kingdom. http://dx.doi.org/10.5258/SOTON/374711

Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill building approach (*5th Ed.). Chichester, U.K.: John Wiley and Sons.

Shakeri, S. (2013). *Data curation perspectives and practices of researchers at Kent State University's Liquid Crystal Institute: A case study* (Doctoral dissertation). Kent State University, USA. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=kent1385382943 (Accessed September 2, 2017).

Sheather, J. (2009). Confidentiality and sharing health information. *BMJ*, 338, b2160. https://doi.org/10.1136/bmj.b2160

Shen, Y. (2016). Research data sharing and reuse practices of academic faculty researchers: A study of the Virginia Tech data landscape. *International Journal of Digital Curation*, *10*(2), 157-175. http://dx.doi.org/10.2218/ijdc.v10i2.359

Silverman, D. (2011). *Interpreting qualitative data: A guide to the principles of qualitative research* (4th Ed.). London Sage Publications.

Soehner, C., Steeves, C., & Ward, J. (2010). *E-science and data support services: A study of ARL member institutions*. Washington, DC: Association of Research Libraries. Retrieved from https://files.eric.ed.gov/fulltext/ED528643.pdf (Accessed September 20, 2017).

Soh, K. (2015). What the overall doesn't tell about world university rankings: Examples from ARWU, QSWUR, and THEWUR in 2013. *Journal of Higher Education Policy and Management*, *37*(3), 295-307. https://doi.org/10.1080/1360080X.2015.1035523

Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management*

Decision, *39*(7), 551-555.

Sutton, R. I., & Staw, B.M. (1995). What theory is not. *Administrative Science Quarterly*, *371*-384. http://dx.doi.org/10.2307/2393788

Swan, A., & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs*. Bristol: JISC Retrieved from http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreportaspx (Accessed January 10, 2017).

Swanson, R.A., & Chermack, T.J. (2013). *Theory building in applied disciplines*. Montgomery Street: Berrett- Koehler Publishers.

Takashima, K., Maru, Y., Mori, S., Mano, H., Noda, T., & Muto, K. (2018). Ethical concerns on sharing genomic data including patients' family members. *BMC Medical Ethics*, *19*(1), 61. https://doi.org/10.1186/s12910-018-0310-5

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, California: Sage Publications.

Tavakol, M., & R. Dennick. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods research*, *1*(1), 77-100. http://dx.doi.org/10.1177/2345678906292430

Teddlie, C., &Tashakkori, A. (2009). Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences. Thousand Oaks, CA: Sage.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS one*, *6*(6), e21101. http://dx.doi.org/10.1371/journal.pone.0021101

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, *36*(2), 84-90. http://dx.doi.org/10.1016/j.lisr.2013.11.003

Teeters, J.L., Harris, K. D., Millman, K. J., Olshausen, B. A., & Sommer, F. T. (2008). Data sharing for computational neuroscience. *Neuroinformatics, 6*(1), 47-55. http://dx.doi.org/1010.1007/s12021-008-9009-y

Thomas, J. (2011). Future-proofing: The academic library's role in e-Research support. *Library Management*, *32*(1), 37-47. http://dx.doi.org/10.1108/01435121111102566

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, *15*(4), 404-428. http://dx.doi.org/10.1177/1534484316671606

Torrance, H. (2012). Triangulation, respondent validation, and democratic participation in mixed methods research. *Journal of Mixed Methods Research, 6*(2), 111-123. http://dx.doi.org/10.1177/1558689812437185

Trochim, W.M., & Donnelly, J.P. (2006). Research methods knowledge base. Retrieved from http://www.anatomyfacts.com/research/researchmethodsknowledgebase.pdf.

United Nations (2014). Malawi country profile. Retrieved from http://www.mw.one.un.org/country-profile/ (Accessed May 23, 2018).

United Nations Development Programme (2016). The 2016 Human Development Report. Retrieved from http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/MWI.pdf (Accessed May 23, 2018).

UNESCO (2014). *Mapping research and innovation in the Republic of Malawi.* UNESCO: Paris.

UNESCO (2017). Global open access portal: Malawi. Retrieved from http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/access-by-region/africa/malawi/ (Accessed 14 January 2018).

UNESCO Institute for Statistics (2019). *Literacy rate, adult total (% of people ages 15 and above).* Retrieved from http://uis.unesco.org/ (Accessed 24 July, 2019).

UNI1 (2017). *UNI1 annual report*. UNI1: Malawi.

UNI2 (2018). *Academic staff records.* UN2: Blantyre, Malawi

University of the South California (2019). Organizing your social sciences research paper: 8. the discussion. Retrieved from http://libguides.usc.edu/writingguide/discussion (Accessed 4 March, 2019).

van Deventer, M., & Pienaar, H. (2015). Research data management in a developing country: a personal journey. *International Journal of Digital Curation, 10*(2), 33-47. http://dx.doi.org/10.2218/ijdc.v10i2.380

van Horn, J.D., & Gazzaniga, M.S. (2013). Why share data? Lessons learned from the fMRIDC. *Neuro-Image,* 82, 677–682. http://dx.doi.org/10.1016/j.neuroimage.2012.11.010

Vardigan, M., & Whiteman, C. (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. *Archival Science, 7*(1), 73-87. https://doi.org/10.1007/s10502-006-9037-z

Vision 2020 (2000). *The national long term development perspective for Malawi*. Lilongwe: National Economic Council.

Wahyuni, D. (2012). The research design maze: understanding paradigms, cases, methods and methodologies. *Journal of Applied Management Accounting Research, 10*(1): 69-80.

Walliman, N. (2006). *Social research methods*. Los Angeles: Sage.

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS one*, *8*(7), e67332. https://doi.org/10.1371/journal.pone.0067332.

Walters, T., & Skinner, K. (2011). *New roles for new times: Digital curation for preservation.* Washington, DC: Association of Research Libraries. Retrieved from http://files.eric.ed.gov/fulltext/ED527702.pdf (Accessed November 12, 2017).

Wassenaar, D. R. (2006). *Ethical issues in social science research*. In: Terre Blanche, M., K. Durrheim and D. Painter (eds). Research in practice: applied methods for the social sciences. University of Cape Town: Cape Town.

Watson, M. (2015). When will 'open science'become simply 'science'?. *Genome biology*, *16*(1), 101. http://dx.doi.org/10.1186/s13059-015-0669-2

Wellington, J.J., Bathmaker, A.M., Hunt, C., McCulloch, G., & Sikes, P. (2005). *Succeeding with your Doctorate*. Thousand Oaks, CA: Sage Publications.

Whitlock, M.C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, *26*(2), 61–65. http://dx.doi.org/10.1016/j.tree.2010.11.006

Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligenc*e 40, 73–76. http://dx.doi.org/10.1016/j.intell.2012.01.004

Wiley, C. (2018). Data sharing and engineering faculty: An analysis of selected publications. *Science & Technology Libraries*. http://dx.doi.org/10.1080/0194262X.2018.1516596

Wilkinson, D., & Birmingham, P. (2003). *Using research instruments: a guide for researchers (*1st Ed.). London: RoutledgeFalmer.

Williamson, K., & Bow, A. (2000). Analysis of quantitative and qualitative data. In: Williamson, K., and Wagga, W., (Eds), *Research methods for students and professionals: information management and systems* (pp: 267-285). Charles Stuart University: New South Wales.

Williamson, G.R. (2005). *Illustrating triangulation in mixed-methodsnursing research*. *Nurse Researcher*, *12*(4), 7-29. http://dx.doi.org/10.7748/nr2005.04.12.4.7.c5955

Willis, J. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Thousand Oaks, CA: Sage.

Whyte, A., & Tedds, J. (2011). *Making the case for research data management*. DCC briefing papers. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/webfm_send/487 (Accessed February 16, 2017).

Winchester, C. L., & Salji, M. (2016). Writing a literature review. *Journal of Clinical Urology*, *9*(5), 308-312. http://dx.doi.org/10.1177/2051415816650133

Wittig, U., Rey, M., Weidemann, A., & Müller, W. (2017). Data management and data enrichment for systems biology projects. *Journal of Biotechnology, 261*, 229-237. http://dx.doi.org/10.1016/j.jbiotec.2017.06.007

Woolfrey, L. (2009). *Archiving social survey data in Africa: an overview of African microdata curation and the role of survey data archives in data management in Africa* (Unpublished doctoral dissertation). University of Cape Town, South Africa.

World Bank & Elsevier (2014). *A decade of development in sub-Saharan African science, technology, engineering and mathematics research*. Washington DC: World Bank

World Bank (2010). *The World Bank working paper no 182: The higher education in Malaw*i. World Bank: Washington DC.

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International Digital Library Perspectives, 23*(4), 335-340. https://doi.org/10.1108/10650750710831466

Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation* (Ddoctoral dissertation). The University of North Carolina at Chapel Hill): USA. Retrieved from https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83 (Accessed September 14, 2017).

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A Content analysis of libraries' websites. *College & Research Libraries, 78*(7), 920-933. https://doi.org/10.5860/crl.78.7.920

Zvyagintseva, L. (2015). *Articulating a vision for community-engaged data curation in the digital humanitie*s (Doctoral thesis). University of Alberta, Canada. https://doi.org/10.7939/R3M66S

LIST OF APPENDICES

Appendix A: Questionnaire library staff

UNIVERSITY *of the*
WESTERN CAPE

QUESTIONNAIRE FOR LIBRARY STAFF

Research data management in public universities in Malawi

SECTION A: BACKGROUND INFORMATION

1. To which institution do you belong?

   UNI2 [ ]
   UNI1 [ ]

2. What is your gender?

   Male [ ]
   Female [ ]

3. What is the highest level of your qualification?

   Certificate [ ]
   Diploma [ ]
   Bachelors [ ]
   Masters [ ]
   PhD [ ]

   Others, please specify ----------

4. Please, indicate your rank/position in the Library

Library Assistant [ ]
Senior Library Assistant [ ]
Assistant Librarian [ ]
Senior Assistant Librarian [ ]
University/College Librarian [ ]

Others, please specify …….

SECTION B: RESEARCH DATA CREATION PRACTICES

5 Do researchers seek help from you in their research activities?

Yes [ ]
No [ ]

6 Which of the following research activities do researchers commonly consult you for help?

Research areas	[ ]
Data collection	[ ]
Data cleaning	[ ]
Data analysis using computer software	[ ]
Data storage and preservation	[ ]
Developing online data collection tools	[ ]
Installation of data analysis software (e.g. SPSS)	[ ]
Recovery of lost research data/information	[ ]
Identification of credible journals	[ ]
Citation and referencing	[ ]
Sources of research collaboration	[ ]
Sources of funding opportunities	[ ]

Others, please specify ----------

SECTION C: PRESERVATION PRACTICES OF RESEARCH DATA

Preservation practices

7 Which of the following digital facilities are available in your library that can be used to preserve research data for researchers? **Select all that apply**

Personal computers	[ ]
Office computers	[ ]
External hard drives	[ ]
CDs for backup	[ ]
Institution's available networked capacity	[ ]
Commercial software or services	[ ]
Freely available software	[ ]
Google Drive	[ ]
Drop Box	[ ]
Flash/USB drive	[ ]
Email account(s)	[ ]

Other, please specify----------

8 For each of the following decisions regarding backing up of research data for researchers, select the best option that represents the extent to which you offer the service to researchers.

Service	I do help them already	I have not helped them before but I am ready to help	I am not ready to help because I lack skills	Not sure
Helping save copies on a local server				
Helping save files on a disk, USB drive, tape, computer hard drive				
Helping them save files on a central campus server				
Helping them save copies on a web-based or cloud server				
Helping them store copies in a data repository or archives				
Restricting access to files				
Others (specify)				

9 For each of the following decisions regarding research data preservation, select the best option that represents the extent to which you may offer your service to researchers.

Types of decisions	I do help them already	I have not helped them before but I am ready to help	I am not ready to help because I lack skills	Not sure
Deciding which data is important to preserve				
Deciding whether data can be safely shared				
Determining standards for de-identifying sensitive data				
Determining what constitutes compliance with commercial licenses, government regulations, funding agency mandates, etc.				
Determining the appropriate metadata to describe data sets (i.e., descriptive information to enable others to reuse data)				
Determining provisions for short-term data storage/preservation (5 years or less)				

10 In your opinion, select the best option that represents your library's readiness in the provision of the following research data preservation services to researchers.

Services	The library is already offering	The library is not offering but it has capability	The library does not have the capability to offer	Not sure
Provision of advanced computing options				
Provision of statistical and other data analysis support				
Short-term data storage/preservation (5 years or less)				

Long-term data storage/preservation (more than 5 years)				
Data security support				
Guidance on depositing data into repositories or archives				
Guidance on how to use appropriate metadata				
Guidance on writing a data management plan				

Technical infrastructure

11 In your opinion, do you think your university provides enough infrastructure to support management of research generated within the university?

Yes [ ]
No [ ]

11.1. If you answered **Yes** to question 11 above, explain the kind of support offered.

11.2. If you answered **No** to question 11, what kind of support would you like your university to provide?

Kind of support	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
Should establish a process for managing data during the life of the project (short-term – 5 years or less)					
Should establish a process for managing data beyond the life of the project (long-term beyond 5 years).					
Should establish necessary tools and technical support for data management during the life of the project (short-term – 5 years or less)					
Should establish necessary tools and technical support for data management data beyond the life of the project (long-term - beyond 5 years).					

Should establish necessary funds to support data management during the life of a research project (short-term -5 years or less)					
Should establish necessary funds to support data management beyond the life of the project (long-term-beyond 5 years).					

SECTION D: COMPETENCY IN DATA CURATION ACTIVITIES

12 Have you ever attended any training workshop(s) on research data management?

Yes [ ]
No [ ]

12.1. If you answered **Yes** to question 12, which of the following organised the workshop?

It was organised my university [ ]
Ii was organised by my university library [ ]
It was organised by a government agency [ ]
It was organised by an international agency [ ]

Others (specify) ----------

12.2. If you answered **No** to question 12, will you be willing to attend such training workshops if an opportunity avails itself and explain the reasons why?

---

13 For each of the following research data management activities, indicate whether you are competent or if you need to be trained by experts?

	I am competent	**I need training**
Preservation planning		
Identifying new standards, practices and software for curation		
Curating and preserving digital objects based the curation lifecycle		
Long term digital data preservation strategies		
Creating preservation metadata (in the library or helping researchers)		
Collecting data from creators, archives, repositories or data centres		

Appraising and selecting digital objects for long term preservation		
Transferring preserved digital objects to strategic storage facilities (e.g. repositories)		
Storing digital information in a secure manner adhering to relevant standards		
Providing access to stored digital objects to bona fide users		
Disposing data not selected for long term preservation		
Repackaging of digital objects		
Migrating digital information to newer file formats that support its continued access and preservation		
Citing, transforming, editing, describing, and sharing data		

SECTION E: CHALLENGES FACED IN THE MANAGEMENT OF RESEARCH DATA

14 To what extent does each of the following limit your involvement in data management activities?

Challenges	**Agree Strongly**	**Agree Somewhat**	**Neutral**	**Disagree Somewhat**	**Disagree Strongly**
Failure by researchers to engage me in data curation					
Lack of incentives to curate data					
There is larger amounts of data to be handled by librarians					
Lack of time to collect data from researchers for curation					
Lack of storage and network infrastructure					
Lack of curation tools and software					
Lack of policy frameworks					
Lack of curation skills and training					
Lack of guidance and support					
Difficulty in finding and accessing data produced by researchers					
Lack of skills to create metadata					
Lack of standardised metadata					
Lack of support from the university in research data management					
Prohibitive institutional policies					
Obsolescence of technologies					
Ethical and legal norms					
Other (Specify):					

Please, feel free to make any comments in relation to research data management in Malawi

---

END OF QUESTIONNAIRE

ONCE AGAIN, THANK YOU FOR TAKING PART IN THIS STUDY

Appendix B: Questionnaire for researchers

UNIVERSITY *of the*
WESTERN CAPE

QUESTIONNAIRE FOR RESEARCHERS

Research data management in public universities in Malawi

SECTION A: BACKGROUND INFORMATION

1. To which institution do you belong?

   UNI2 [ ]
   UNI1 [ ]

2. What is your gender?

   Male [ ]
   Female [ ]

3. To which Faculty do you belong?

---

4. To which Department or Centre do you belong?

---

5. What is your highest qualification?
   Masters [ ]
   PhD [ ]
   Post PhD [ ]

   Others (Please specify) ----------

6. What is your current rank in the university?
   Lecturer [ ]
   Senior Lecturer [ ]
   Associate Professor [ ]
   Professor [ ]
   Senior Professor [ ]

   Others (Please specify) ----------

SECTION B: RESEARCH DATA CREATION, SHARING AND RE-USE PRACTICES

Data creation practices

7. Which of the following best describes your research out for the past 10 years?

Research activities	**Research output**				
	1-5	**6-10**	**11-15**	**16-20**	**21 and above**
Number of papers already published in peer reviewed journals					
Number of papers currently in review					
Number of research projects in progress					
Number of commissioned reports (completed or in progress)					
Others (Specify)					

8 Which of the following data format(s) do you generate through your research process? **Please select all that apply**

Digital text or digital copies of text [ ]
Digital images or digital copies of images [ ]
Audio recordings [ ]
Video recordings [ ]
Spreadsheets [ ]
Digital databases (e.g., surveys, census data, government statistics, etc.) [ ]
Computer code [ ]
Biological/organic/inorganic samples or specimens [ ]
Spatial data [ ]
Artistic products [ ]

Others (Specify): ----------

Data sharing practices

9 Do you usually share your research data with other researchers or stakeholders?

Yes [ ]
No [ ]

9.1. If you answered **Yes** to question 9, which of the followings factors motivate or compel you to share your research data? **Select all that apply.**

Journal policies require me to submit my manuscripts with data [ ]
Research funders compel me to share data from research projects they have funded [ ]
My university requires me to share the data from my research projects [ ]
I share data because open access proponents have convinced me to do so [ ]
I share data because I personally find it scientifically necessary [ ]

9.2.If you answered **Yes** to question 9, which of the following best represent the ways you share your data? **Please, select all that apply.**

Sharing practices	All	Most	Some	None
Through external drives (flask disks)				
Through emails				
Through e-journals' websites				
On social networks				
On my personal website/blogs/wikis				
Through clouds (Google Drive, DropBox, etc)				
University repositories				
Through research funders website				
On my university's website				
On the principal investigator's website				
Through a national network				
Through a regional network				
Through a global network				
Other (Specify)				

10 For each of the following factors, indicate the extent to which they discourage you from sharing your research data with other researchers.

Statement	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Lack of incentives					
Lack of funding					
Lack of standards or guidelines for sharing data					
The data is not fully documented					
There is no place to put the data					
License agreements prohibit sharing data					
I would lose control over my data					
I have insufficient skills to make my data available to the public					
The data is in a format that is not widely readable					
My data may be misinterpreted by others					
The university owns the data I produce					
If funded, the funding agency owns the data					
Insufficient time					

11 For each of the following conditions, indicate the extent to which they could encourage you to share your research data with others.

Conditions	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
I would be willing to place at least some of my data into a central data repository with no restrictions					
I would be willing to place all of my data into a central data repository with no restrictions					
I would be more likely to make my data available if I could place conditions on access					
I would be willing to share data across a broad group of researchers who use data in different ways					
It is important that my data are cited when used by other researchers.					
It is appropriate to create new datasets from shared data					
Others (Specify)					

Research data re-use practices

12 How frequently do you use research data produced/created by other researchers or research institutions in your research activities?

Always [ ]
Frequently [ ]
Occasionally [ ]
Seldom [ ]
Never [ ]

13 For each of the following factors, indicate the extent to which they discourage you from using research data produced by other researchers or research institutions.

Factors	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Difficult to find, discover, or access reusable data					
Hard to integrate with my own data					

Not trusting others' collection methods					
Data may be misinterpreted due to complexity of the data					
Lack of common or standard formats					
Lack of adequate metadata/data description information					
Data may be misinterpreted due to poor quality of the data					
Data may be used in other ways than intended					
Legal/ethical restrictions					
Other (specify)					

SECTION C: PRESERVATION PRACTICES OF RESEARCH DATA

Preservation practices

14 Do you think it is necessary to preserve your research data?
Yes [ ]
No [ ]

14.1. If you answered **Yes** to question14, for how long do you think your data will remain valuable?

Indefinitely [ ]
10 – 20 years [ ]
5–10 years [ ]
3–5 years [ ]
1-2 [ ]
Not sure [ ]

15 What is the largest amount of digital research data for a single research project you have worked on in the past?
1 GB (gigabyte) or less [ ]
More than 1 GB but less than 100 GB [ ]
More than 100 GB but less than 1 TB (terabyte) [ ]
More than 1 TB but less than 100 TB [ ]
More than 100 TB but less than 1 PB (petabyte) [ ]
More than 1PB [ ]
I don't know [ ]

16 Which of the following digital facilities do you use to store your data? **Select all that apply**

Personal computers	[ ]
Office computers	[ ]
External hard drives	[ ]
CDs for backup.	[ ]
Institution's available networked capacity	[ ]
Commercial software or services	[ ]
Freely available software or services (Google Drive)	[ ]
Flash/USB drive	[ ]
Email account(s)	[ ]

Others (Specify)----------

17 Have you put in place some strategies to protect your data from loss?

Yes [ ] No [ ]

17.1. If you answered **No** to question 17, please provide the reasons?

17.2. If you answered **Yes** to question 17, which of the following strategies have you adopted to protect your data from loss? **Please select all that apply**

Copies are uploaded on Goodge Drive	[ ]
Copies are uploaded on Drops Box	[ ]
Copies are kept in my email	[ ]
Copies of data sets are saved on a disk, USB drive, tape, computer hard drive	[ ]
Copies of data sets are saved on a local server	[ ]
Copies of data sets are saved on a central campus server	[ ]
Copies of data sets are saved on a web-based or cloud server	[ ]
Copies of data sets are stored in a data repository or archives	[ ]
Backup files are automatically generated	[ ]
Backup files are manually generated	[ ]

Others (Specify): ----------

Technical infrastructure

18 In your opinion, do you think your university provides enough infrastructure to support your research data management?

Yes [ ]
No [ ]

18.1. If you answered **Yes** to question 18 above, explain the kind of support offered.

18.2. If you answered **No** to question 18, to which extent do each of the following kind of support would you like your university to provide?

Kind of support	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Should establish a process for managing data during the life of the project (short-term – 5 years or less)					
Should establish a process for managing data beyond the life of the project (long-term beyond 5 years).					
Should establish necessary tools and technical support for data management during the life of the project (short-term – 5 years or less)					
Should establish necessary tools and technical support for data management data beyond the life of the project (long-term - beyond 5 years).					
Should establish necessary funds to support data management during the life of a research project (short-term -5 years or less)					
Should establish necessary funds to support data management beyond the life of the project (long-term-beyond 5 years).					

SECTION D: COMPETENCIES REQUIRED FOR RESEARCH DATA MANAGEMENT

19 Have you ever attended any training workshop(s) on research data management?

Yes [ ] No [ ]

19.1 If you answered **yes** to question 19, which of the following organised the workshop? **Select all that apply**

It was organised by my university [ ]
Ii was organised by my university's library [ ]
It was organised by a government agency [ ]
It was organised by an international agency [ ]

Others (specify) ----------

19.2 If you answered **No** to question 19, will you be willing to attend such training workshops if an opportunity avails itself? Explain your reasons.

---

20 Do you assign metadata (description of data) to your research data?

Yes [ ]

No [ ]

20.1. If you answered **Yes** to question 20, which of the following metadata do you use in describing your data? **Select all that apply.**

No metadata standard [ ]
Metadata standardised within my lab [ ]
International Standards Organisation (ISO) [ ]
Open GIS [ ]
Ecological Metadata Language [ ]
Federal Geographic Data Committee [ ]
Dublin Core [ ]
Darwin Core [ ]
Directory Interchange Format [ ]

Others, (please specify) ----------

21 For each of the following research data management activities, indicate whether you are competent or if you need to be trained by experts?

Curation activities	**I am competent**	**I need training**
Developing and writing a data management plan		
Advanced computing options		
Short term digital data preservation strategies		
Long term digital data preservation strategies		
Preservation planning		
Identifying new standards, practices and software for curation		
Creating preservation metadata for describing my data sets		
Guidance on depositing data into repositories or archives		
Storing digital information in a secure manner adhering to relevant standards		
Disposing data not selected for long term preservation		

Migrating digital information to newer file formats that support its continued access and preservation		

22 How often do you seek help on managing your research data from the following professionals?

Professionals	Always	Frequently	Occasionally	Not sure	Never
Librarians					
ICT experts					
Fellow researchers					
Research director					
Others (please, specify)					

SECTION E: CHALLENGES IN THE MANAGEMENT OF RESEARCH DATA

23 How frequently do you lose your research data due to the following?

Factor	Always	Frequently	Occasionally	Seldom	Never
Obsolescence of technologies					
Accidental damage of storage facilities					
Stolen storage facilities (flash discs, laptops, etc.)					
Others (please, specify) ……					

24 To what extent does each of the following present any challenges with regard to your research data management and re-use?

Challenges	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Lack of incentives to share data					
Lack of storage and network infrastructure					
Lack of curation tools and software					
Lack of policy frameworks					
Lack of curation skills and training					
Lack of guidance and support					
Difficulty in finding and accessing data produced by other researchers					
Most data is not trustworthy					

Lack of skills in sharing my data with other researchers					
Tracking updates to data (i.e., versioning)					
Lack of skills to create metadata					
Lack of standardised metadata					
Failure by data re-users to cite the data I generated.					
Lack of support from the university in research data management					
Prohibitive institutional policies					
Obsolescence of technologies					
Ethical and legal norms					
Other (Specify):					

Please, feel free to make any comments in relation to research data management at your university.

END OF QUESTIONNAIRE

THANK YOU FOR TAKING PART IN THIS STUDY

Appendix C: Interview guide for directors of research

UNIVERSITY *of the*
WESTERN CAPE

INTERVIEW GUIDE FOR DIRECTORS OF RESEARCH

Research data management in public universities in Malawi

SECTION A: PERSONAL INFORMATION

- How long have you been working at this university?
- How important is research in your professional life?
- What are the duties attached to your office?

SECTION B: RESEARCH DATA CREATION, SHARING AND RE-USE PRACTICES

Creation practices

- How does the university encourage lecturers to conduct research and publish in creditable journals?
- Think of research at this university as one of the core functions of the university. What has been the role of your office in promoting research output?
- What do you understand by digital research data?
- According to the university policy, who owns the research data produced by lecturers employed by the university in terms of self-sponsored research, university funded research and externally funded research?

Sharing practices

- What role does your office play to encourage sharing of research data within the university?
- What are the key ways of sharing research data and findings within the university?
- For research projects that the university funds, do you need researchers to provide you with data in addition to the actual findings? Is this data shared with other researchers?

- How frequently do external funders require your university research teams to provide them with data in addition to the actual results?
- Does your university have a policy on research data sharing? What are the requirements of the policy?
- Why do you think it is necessary for researchers to share their research data in addition to research findings?

Re-use practices

- How does your office view the concept of research data re-use?
- Do researchers use data created or generated by other researchers or research institutions?
- Does the university encourage researchers to use research data or generated by other researchers or research institution?

SECTION C: PRESERVATION PRACTICES OF RESEARCH DATA

Preservation practices

- What are the storage facilities that the university provides its researchers for storage of digital research data?
- What measures has the university put in place to ensure that digital research data is properly preserved for longevity?
- What role does your office play to ensure that researchers, librarians and ICT staff work together in managing research data?

Technical infrastructure

- How is your university building research data management infrastructure to ensure research data is preserved?

SECTION D: COMPETENCIES IN DIGITAL CURATION ACTIVITIES

- Explain how your office or the university supports lecturers and data curators in research data management activities.

- **SECTION E: FACTORS AFFECTING MANAGEMENT OF RESEARCH DATA**
  - What factors affect research data management by lecturers in your university?

If you have additional comments in relation to the topic under discussion, please feel free to do so.

End of interview. Please feel free to make any comments in relation to the topic we have discussed.

Appendix D: Information consent letter for librarians

UNIVERSITY *of the*
WESTERN CAPE

Department of Library and Information Science

Private Bag X17, Bellville, 7535, Cape Town, South Africa

INFORMATION CONSENT LETTER: LIBRARIANS

Dear Sir/ Madam,

My name is Winner Chawinga. I am a doctoral student from the Department of Library and Information Science at the University of the Western Cape. I am conducting a survey amongst researchers, librarians and research directors from UNI1 and UNI2 to understand research data management practices in these Malawian public universities. The survey is for the accomplishment of my doctoral thesis.

The topic of my research project is **"Research data management in public universities in Malawi."** The objectives of my research are as follows:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers need to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affect the management of research data in public universities in Malawi

I am therefore requesting you in your capacity as a library professional to please participate in this survey. Enclosed is a questionnaire that takes a variety of questions about the topic under study.

No names are required and your identity will remain anonymous. If you agree to participate, please read and confirm your participation by ticking the consent box. As the consent form

indicates, your participation is completely voluntary, your identity remains anonymous, your responses will be kept confidential, and you are free to withdraw from the study at any time.

If you have any questions or concerns or wish to know more about this study, please contact me, Winner Chawinga at 3371323@myuwc.ac.za, or you could contact my supervisor Prof Sandy Zinn at szinn@uwc.ac.za. Your participation in this study is greatly appreciated.

Yours Sincerely,

Winner Chawinga (PhD candidate)
Department of Library & Information Science
Faculty of Arts
University of the Western Cape
P/Bag X17
Bellville 7535
Phone; +265993509295
Email: 3371323@myuwc.ac.za or winnchawinga@gmail.com

Appendix E: Information consent letter for researchers

UNIVERSITY *of the*
WESTERN CAPE

Department of Library and Information Science

Private Bag X17, Bellville, 7535, Cape Town, South Africa

INFORMATION CONSENT LETTER: RESEARCHERS

Dear Sir/ Madam,

My name is Winner Chawinga. I am a doctoral student from the Department of Library and Information Science at the University of the Western Cape. I am conducting a survey amongst researchers, librarians and research directors from UNI1 and UNI2 to understand research data management practices in these Malawian public universities. The survey is for the accomplishment of my doctoral thesis.

The topic of my research project is **"Research data management in public universities in Malawi''**

The objectives of my research are as follows:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers need to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affect the management of research data in public universities in Malawi

I am therefore requesting you in your capacity as a lecturer or researcher to please participate in this survey. Enclosed is a questionnaire that takes a variety of questions about the topic under study. No names are required and your identity will remain anonymous. If you agree to participate, please read and confirm your participation by ticking the consent box. As the consent form indicates, your participation is completely voluntary, your identity remains

anonymous, your responses will be kept confidential, and you are free to withdraw from the study at any time.

If you have any questions or concerns or wish to know more about this study, please contact me, Winner Chawinga at 3371323@myuwc.ac.za, or you could contact my supervisor Prof Sandy Zinn at szinn@uwc.ac.za. Your participation in this study is greatly appreciated.

Yours Sincerely,

Winner Chawinga (PhD candidate)

Department of Library & Information Science
Faculty of Arts
University of the Western Cape
P/Bag X17
Bellville 7535
Phone; +265993509295
Email: 3371323@myuwc.ac.za or winnchawinga@gmail.com

Appendix F: Information consent letter for directors of research

UNIVERSITY *of the*
WESTERN CAPE

Department of Library and Information Science

Private Bag X17, Bellville, 7535, Cape Town, South Africa

INFORMATION CONSENT LETTER: DIRECTORS OF RESEARCH

Dear Sir/ Madam,

My name is Winner Chawinga. I am a doctoral student from the Department of Library and Information Science at the University of the Western Cape. I am conducting a survey amongst researchers, librarians and research directors from UNI1 and UNI2 to understand research data management practices in these Malawian public universities. The survey is for the accomplishment of my doctoral thesis.

Aim of study and objectives

I want to interview you to collect data as part of my PhD research at University of the Western Cape. The study is about understanding research data management in universities in Malawi. The topic of my research project is **"Research data management in public universities in Malawi''** The objectives of my research are as follows:

- To determine research data creation, sharing and re-use practices in public universities in Malawi;
- To investigate research data preservation practices in public universities in Malawi;
- To investigate competencies that librarians and researchers need to effectively manage research data in public universities in Malawi; and
- To find out the challenges that affect the management of research data in public universities in Malawi

Estimated duration of the interview

This interview will last between 30 and 60 minutes.

Voluntary participation and confidentiality

Your participation in this study is voluntary and you may wish to terminate the interview or refuse to answer any question at any stage during the interview. No part of our conversation will be attributable to you. The anonymous conversation will be digitally recorded, transcribed and coded in order for me to develop themes and categories for moving to the next stage of my research. I will also occasionally be taking down notes which would help me when I am listening to the audio recording afterwards. Should you inadvertently mention any names of individuals or reveal any particularly sensitive information during the interview, I will remove them from the transcribed text so as to protect confidentiality. I will provide you with a copy of the transcribed conversation for you to sense check and return with amendments if you so wish.

I am therefore requesting you to please participate in this study. If you agree to participate, please read and confirm your participation by ticking the consent box. As the consent form indicates, your participation is completely voluntary, your identity remains anonymous, your responses will be kept confidential, and you are free to withdraw from the study at any time.

Further information

If you have any questions or concerns or wish to know more about this study, please contact me, Winner Chawinga at 3371323@myuwc.ac.za, or you could contact my supervisor Prof Sandy Zinn at szinn@uwc.ac.za. Your participation in this study is greatly appreciated

Yours Sincerely,

Winner Chawinga (PhD candidate)
Department of Library & Information Science
Faculty of Arts
University of the Western Cape
P/Bag X17
Bellville 7535
Phone; +265993509295
Email: 3371323@myuwc.ac.za or winnchawinga@gmail.com

Appendix G: Ethics clearance

UNIVERSITY of the WESTERN CAPE

OFFICE OF THE DIRECTOR: RESEARCH
RESEARCH AND INNOVATION DIVISION

Private Bag X17, Bellville 7535
South Africa
T: +27 21 959 4111/2948
F: +27 21 959 3170
E: research-ethics@uwc.ac.za
www.uwc.ac.za

31 July 2018

Mr W Chawinga
Library and Information Science
Faculty of Arts

Ethics Reference Number: HS18/4/16

Project Title: Research data management in public universities in Malawi: an interdisciplinary study of medical sciences and humanities.

Approval Period: 27 July 2018 – 27 July 2019

I hereby certify that the Humanities and Social Science Research Ethics Committee of the University of the Western Cape approved the methodology and ethics of the above mentioned research project.

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

Please remember to submit a progress report in good time for annual renewal.

The Committee must be informed of any serious adverse event and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

PROVISIONAL REC NUMBER - 130416-049

FROM HOPE TO ACTION THROUGH KNOWLEDGE.

Appendix H: Request to undertake research at one of the two universities

UNIVERSITY of the WESTERN CAPE

Faculty of Arts

Department of Library & Information Science

March 14, 2018

The Registrar

Dear Sir or Madam,

Request to conduct a study

This letter serves to introduce Mr. Winner Chawinga who is a PhD student in the Department of Library and Information Science at the University of the Western Cape in South Africa.

Mr. Chawinga's research topic is: *Research data management in public universities in Malawi* His particular interest is to holistically investigate how research data is generated, organised, shared, stored and preserved in medical sciences and humanities for the purpose of long-term access and re-use. The study will include distributing a questionnaire and conducting interviews with some stakeholders in public universities in Malawi The issues that Mr. Chawinga is researching are both important and timely in this research area. To this end, I request your good Office to grant Mr. Winner Chawinga permission to conduct this study at

Mr. Chawinga's PhD proposal is ready for submission to the University of Western Cape's Humanities and Social Sciences Research Ethics Committee for ethical clearance but it can only be submitted if permission to conduct a study at your university is granted. Mr. Chawinga's research proposal will be scrutinised and approved by the University of the Western Cape Ethics Committee and I assure you that all research ethical issues will be observed.

I would like to thank you in advance for your assistance.

Yours sincerely

S. E. Zinn

Prof Sandy Zinn
Chairperson, Department of Library & Information Science,
Supervisor for Mr. Chawinga

HOD: Prof Sandy Zinn, szinn@uwc.ac.za
Private Bag X17 Bellville 7535 South Africa
T: +27 (0)21 959 2349/2137
www.uwc.ac.za/arts/

Tel: +27 21 959 4111
Email: research-ethics@uwc.ac.za

A place of quality, a place to grow, from hope to action through knowledge

Appendix I: Request to undertake research at one of the two universities

UNIVERSITY of the WESTERN CAPE

Faculty of Arts

Department of Library & Information Science

March 14, 2018

The Registrar

Dear Sir or Madam,

Request to conduct a study

This letter serves to introduce Mr. Winner Chawinga who is a PhD student in the Department of Library and Information Science at the University of the Western Cape in South Africa.

Mr. Chawinga's research topic is: *Research data management in public universities in Malawi* His particular interest is to holistically investigate how research data is generated, organised, shared, stored and preserved in medical sciences and humanities for the purpose of long-term access and re-use. The study will include distributing a questionnaire and conducting interviews with some stakeholders in public universities in Malawi The issues that Mr. Chawinga is researching are both important and timely in this research area. To this end, I request your good Office to grant Mr. Winner Chawinga permission to conduct this study at

Mr. Chawinga's PhD proposal is ready for submission to the University of Western Cape's Humanities and Social Sciences Research Ethics Committee for ethical clearance but it can only be submitted if permission to conduct a study at your university is granted. Mr. Chawinga's research proposal will be scrutinised and approved by the University of the Western Cape Ethics Committee and I assure you that all research ethical issues will be observed.

I would like to thank you in advance for your assistance.

Yours sincerely

J. E Zinn

Prof Sandy Zinn
Chairperson, Department of Library & Information Science,
Supervisor for Mr. Chawinga

HOD: Prof Sandy Zinn, szinn@uwc.ac.za
Private Bag X17 Bellville 7535 South Africa
T: +27 (0)21 959 2349/2137
www.uwc.ac.za/arts/

Tel: +27 21 959 4111
Email: research-ethics@uwc.ac.za

A place of quality, a place to grow, from hope to action through knowledge

Appendix J: Permission to undertake research at one of the two universities

MZUZU UNIVERSITY

OFFICE OF THE UNIVERSITY REGISTRAR

Private Bag 201
Luwinga
Mzuzu 2
MALAWI
Tel.: (265) 01 320 722/575
Fax: (265) 01 320 505
E-mail: ur@mzuni.ac.mw

Ref: MU/1/P1.09 **19th March 2018**

Mr. Winner Chawinga
University of the Western Cape
Department of Library and Information Science
P/Bag X17
South Africa
Bellville 7535

3371323@myuwc.ac.za /winnchawinga@gmail.com

Dear Mr. Chawinga,

APPROVAL TO CONDUCT A STUDY AT MZUZU UNIVERSITY

I refer to your letter dated 14th March, 2018 in which you are requesting for approval to conduct a PhD research at Mzuzu University. I am pleased to inform you that your request has been accepted based on your research title of *research data management in public universities in Malawi*

This permission is granted on the understanding that the information that will be collected will be strictly used for academic purposes.

I wish you well in your studies.

Yours sincerely,

OFFICE OF THE UNIVERSITY REGISTRAR
MZUZU UNIVERSITY
19 MAR 2018
PRIVATE BAG 201

JAMES KWILIMBE
ACTING UNIVERSITY REGISTRAR

Appendix K: Permission to undertake research at one of the two universities

COLLEGE OF MEDICINE

Principal

M. H. C. Mipando MSc, PhD

Our Ref.:

Your Ref.:

College of Medicine
Private Bag 360
Chichiri
Blantyre 3
Malawi
Telephone: 01 871 911
01 874107
Fax: 01 874 700

Dear Winner Chawenga,

REQUEST TO CONDUCT STUDY AT COLLEGE OF MEDICINE

I refer to your letter dated 14th March, 2018 in which you made a request for approval to conduct a study at College of Medicine and the research is in fulfilment for the award of Doctor of Philosophy in Library and Information Science at University of Western Cape, South Africa.

I write to inform you that your request to conduct a study has been approved.

Orama Lipenga

For Registrar