DEVELOPING AN IMPLEMENTATION PLAN FOR RESEARCH DATA MANAGEMENT (RDM) AT THE UNIVERSITY OF GHANA (UG)

Mini-dissertation

by

Bright Kwaku Avuglah

(u15353402)

Submitted in partial fulfilment of the requirements for the degree

MASTER OF INFORMATION TECHNOLOGY

in the

SCHOOL OF INFORMATION TECHNOLOGY

of the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

Supervisor: Emeritus Prof. Peter G. Underwood

December 2016

DECLARATION

I hereby declare:

1. This dissertation is my own original work.
2. Opinions and conclusions drawn in this study are solely mine, and where the works or ideas of others were used by way of direct quotation, paraphrasing or summarisation, these have been duly acknowledged through in-text citations and a complete list of references at the end of the report.
3. This research work has not been already submitted to any other institution or university for the award of a degree.

30th December, 2016

…….. ……..

STUDENT'S SIGNATURE DATE

(Mr. B.K. Avuglah)

…….. ……..

SUPERVISOR'S SIGNATURE DATE

(Emeritus Prof. P. G. Underwood)

DEDICATION

I dedicate this work to my dear family, the Avuglah/Nyasor family:

Francis (my dad), Ellen (my mom), Janet (my step-mom) and my siblings:

Reuben

Diana

Justice

Daniel

David

Gertrude

Albert

Bismark

ACKNOWLEDGEMENTS

The exercise of writing this dissertation has been the single most challenging experience yet of my life. But He that is El Shaddai sustained me. What can I say?

"Ebenezer, Thus far the LORD has helped me"

Now, I take this opportunity to appreciate all those who in various ways contributed to this success story. To begin with, I am deeply grateful to the Carnegie Corporation of New York, for your generous funding and to the University of Pretoria, South Africa for the wonderful opportunity granted me to participate in the MIT programme.

I owe a special thanks to Emeritus Prof. Peter G. Underwood for dedicating your time and effort to guide me through this research work despite your undoubtedly busy schedule and workload. Your insightful comments and critical questions have helped to shape my perspective on the topic. Your patience, understanding and sympathy towards me have helped me to escape perhaps what would have been an excruciating fate.

Next, to my former boss, Mr. Emmanuel Darkey, God richly bless you, Sir, for your kindness, understanding, support, and encouragement throughout my study. To my family, I want to express my profound gratitude to you all for enduring what looked like total neglect just so I can complete this study.

I also owe a debt of gratitude to all the respondents who agreed to participate in my study and for making time out of your busy schedules to grant me interviews. I am deeply grateful.

Two of my wonderful lecturers, Dr. Martie van Deventer, CSIR, South Africa and Dr. Heila Pienaar, UP, South Africa, both deserve a special mention for willingly answering my questions and clarifying issues that were puzzling to me when I called on them. I also appreciate the efforts of the following persons towards the successful completion of this research work: Prof. Andrew A. Adjei, Director, Office of Research, Innovation and Development (ORID); Mr. Samuel Amebley, Assistant Registrar, Balme Library and Mr. Moses Bediako, Assistant Librarian, Balme Library.

Finally, THANK YOU to my friends and MIT5 colleagues, Evelyn M. Anduvare and Paul Ashaley Nikoi. Both of you in many ways kept me going when it got tough and unbearable. With you, I could still laugh out loud even when I was stressed and helplessly anxious. With you, I had friends who had the right words of encouragement when I wanted to give up. You did all these while working on your own dissertations and had your own stresses to deal with.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AIDA	Assessing Institutional Digital Assets
ANDS	Australian National Data Service
ARC	Australian Research Council
ARL	Association of Research Libraries
AVCC	Australian Vice-Chancellors' Committee
BIS	Department for Business, Innovation & Skills
BRISSKit	Biomedical research infrastructure software service kit
CARDIO	Collaborative Assessment of Research Data Infrastructure and Objectives
CODATA	Committee on Data for Science and Technology
COP	Community of Practice
CPD	Continuing Professional Development
CSIR	Council for Scientific and Industrial Research
DaaS	Database as a Service
DAF	Data Asset Framework
DAFD	Data Audit Framework Development
Damaro	Data Management Roll-out at Oxford
DataNet	Sustainable Digital Data Preservation and Access Network Partners
DataONE	Data Observation Network for Earth
DC	Data Conservancy
DCC	Digital Curation Centre
DCR	Direct Cost Recovery
DIRISA	Data Intensive Research Initiative of South Africa
DMP	Data Management Plan
DOI	Digital Object Identifier
DRAMBORA	Digital Repository Audit Method Based On Risk Assessment
EIDCSR	Embedding Institutional Data Curation Services in Research
EPSRC	Engineering and Physical Sciences Research Council

ESRC	Economic and Social Research Council
FATPOU	Free at the point of use
fEC	Full Economic Costing
HATII	Humanities Advanced Technology and information Institute
HEFCE	Higher Education Funding Council for England
HEI	Higher Education Institution
HKUST	Hong Kong University of Science and Technology
HPC	High Performance Computing
IaaS	Infrastructure as a Service
ICT	Information and communication technology
IDMP	Integrated Data Management Planning
IT	Information Technology
JISC	Joint Information Systems Committee
JISCMRD	Joint Information Systems Committee Managing Research Data
LIASA	Library and Information Association of South Africa
LIS	Library and Information Science
LISC	Library and Information Studies Centre
MaDAM	Manchester Data Management
MANTRA	Research Data Management Training
NCRIS	National Collaborative Research Infrastructure Strategy
NeDICC	Network of Data and Information Curation Community
NERC	Natural Environment Research Council
NHMRC	National Health and Medical Research Council
NICIS	National Integrated Cyber-infrastructure System
NIH	National Institute of Health
NRF	National Research Foundation
NSF	National Science Foundation
OECD	Organisation for Economic Co-operation and Development

OMERO	Open Microscopy Environment Remote Objects
ORID	Office of Research, Innovation and Development
OSTP	Office of Science and Technology Policy
RCUK	Research Council UK
RDM	Research Data Management
RDS	Australia's Research Data Service
RDSF	Research Data Storage Facility
RIN	Research Information Network
RLUK	Research Libraries UK
RUGIT	Russel Group of IT Directors
SAC	Storage, Archive and Curation
SANReN	South African National Research Network
Sudamih	Supporting Data Management Infrastructure for the Humanities
TRAC	Transparent Approach to Costing
UC3	University of California Curation Center
UCT	University of Cape Town
UG	University of Ghana
UGCS	University of Ghana Computing Systems
UGLS	University of Ghana Library Systems
UK	United Kingdom
UKRDS	UK Research Data Service
UMF	University Modernisation Fund
UNISA	University of South Africa
UP	University of Pretoria
USA	Unites States of America
UWE	University of the West of England
VCUL	Virginia Commonwealth University Library
VRE	Virtual Research Environment

ABSTRACT

The current global and data intensive outlook of research provides new opportunities and challenges for HEIs including effective and sustainable RDM. As a growing area of interest in the global research arena, experiences from developed countries have dominated the body of literature on RDM. This study is in part, to fill this gap by assessing the state of the art of RDM and institutional preparedness at the University of Ghana (through existing data management activities and capabilities) in order to develop a plan for implementation.

The study used a qualitative case study method and gathered data using semi-structured interviews and document analysis. Thematic analysis method was used to analyse the data collected. A total of seven respondents (five service providers and two senior researchers) were selected purposively using two sampling techniques ("priori criteria sampling" and snowball sampling). Criteria were set for their inclusion and each respondent provided information about institutional support, capabilities, policies and expectations on RDM.

The findings of the study revealed a number of RDM related activities, these include support for collaborative research, support for data analysis and computational science, guidance on RDM and grant applications as well as support for storage and high-speed connectivity to facility the research enterprise at UG. In terms of capabilities, no specific RDM policy was identified, existing infrastructure identified include an HPC cluster, a private cloud facility (HP Cloud Matrix), an Institutional repository (UGSpace), an institutional Google Drive platform, data analysis packages (NVivo and SPSS) and a robust network and security infrastructure. These were not necessarily provisioned for RDM purposes. Also, the findings show that staff do not possess the necessary skills or adequate knowledge to fully support RDM at UG.

In terms of the specific objectives of the study, the results of the semi-structured interviews and document analysis provided an understanding of the current situation (i.e. requirements, current activities and capabilities at the UG) which is the first objective of the study. These findings were then benchmarked against the EPSRC policy framework following the outline of the DCC CARDIO Matrix and using the optimal desirable expectation or level of development as the standard for comparison. This was useful in identifying gaps in RDM awareness, support and capabilities at UG which is the second objective of the study. To achieve the third objective, which was identifying priority areas for RDM development, the researcher examined both initial findings (i.e. findings on requirements, current activities and capabilities identified under the first objective as well as the gaps identified in the second objective) and proposed six broad areas where UG must focus its RDM development agenda. Finally, the six broad areas proposed in objective three were further cascaded into a number of specific initiatives and tasks to be implemented. This was done taking cognisance of the potential of current infrastructure, gaps identified in institutional awareness and capabilities as well as essentials for a cultural changed.

The study concluded that RDM at the University of Ghana is currently underdeveloped but with immense potential for growth. While a few RDM related activities were identified, existing capabilities were generally found to be inchoate, uncoordinated and not formally instituted.

The study recommended six main areas where the UG should focus RDM development, these include: constituting a steering group to spearhead and coordinate RDM development at the UG, developing a coordinated policy framework for RDM at UG, streamlining existing technical infrastructure to support data management requirements, creating opportunities for RDM training and capacity development for professional staff, researchers and students, developing services to support requirements, and exploring internal funding strategies to facilitate RDM development and support at the UG. The study also recommends that the academic community at the UG should be actively engaged throughout the RDM development process as this is critical to ensure that the eventual solutions are fit for purpose and acceptable.

Keywords: Research Data Management; Research data; RDM implementation; RDM development; RDM Service; University of Ghana, RDM roadmap, RDM implementation plan.

CHAPTER ONE: INTRODUCTION

1.1. Background to the study

The discussions on the subject of Research Data Management (RDM) have mostly concentrated in the developed world. Popular cases in the UK, USA and Australia are often cited (Kahn et al., 2014:296; Jones, Pryor & Whyte, 2013:2; Delserone, 2008:202-203). RDM essentially encapsulates all the activities that capture, document, organise, curate, preserve and provide continued access to research data during and after the research process.

Data is a critical element in any scientific inquiry which forms a basis for generating new knowledge and evidence for verifying and validating research outcomes (Jones, Pryor & Whyte, 2013:1-2; Lynch & Carleton, 2009:236). For many higher education institutions (HEIs), research data represent a valuable institutional asset, which has the potential to be reused or repurposed beyond the active research phase (Jones, Pryor & Whyte, 2013:2-3). This is true, first, because HEIs underwrite a significant number of the research projects conducted by their faculty members, most of whom typically do not get external funding for their research (Lynch & Carleton, 2009:237) and therefore have a stake in the ownership and stewardship of the outputs (including research data) emanating from such "unfunded" research. Secondly, it has been widely acknowledged and eloquently demonstrated how the sharing, reuse and repurposing of research data can advance scientific research and also contribute to industry and humanity at large, as can be found in research areas like climate change and genome sequencing (Holdren, 2013; Harris-Pierce & Liu, 2012; Lynch & Carleton, 2009). Lastly, the economic implications of wringing out the full benefits of research data are enormous especially because of the volatility that affected government subventions to many of these HEIs.

What is more, many research funding bodies – not just those affiliated to government but also private foundations – are becoming more interested in the research data as a valuable output of the research enterprises they fund. The National Science Foundation (NSF) [USA], National Institute of Health (NIH) [USA], Engineering and Physical Sciences Research Council (EPSRC) [UK] and Australian Research Council (ARC) among many others either have guidelines for or require researchers they fund to demonstrate some commitments to data management and data sharing by presenting an RDM plan which will usually capture institutional and researcher responsibilities toward data generated in the course of doing research.

Beyond the institutional benefits and funder expectations, good data management and sharing practices have far-reaching benefits for individual researchers and the broader scientific community. Among the several benefits and reasons documented are: long-term discoverability of data; use and re-use; quality research output which is critical for researcher visibility and research impact (increased citation); easy and secured data sharing across distributed institutional research teams; a growing requirement for assessing

research grants; protection against loss of valuable data or data entropy; and compliance with industry requirements and etiquette (Kahn et al., 2014:297; Jones, Pryor & Whyte, 2013:2; Strasser et al., 2012:1-2; Pienaar, 2011:10).

In Africa, concern for RDM is developing. In the case of South Africa, which is by far the most significant example on the continent, the National Research Foundation's (NRF) recognition of data sharing in its Statement on Open Access to Research Publications and the development of a number of cyber-infrastructure initiatives to support collaborative research is indicative of a national level support. At the institutional level, many HEIs in responding to this recognition are beginning to develop data management policies and programmes (Kahn et al., 2014:296). In a nutshell, issues of data management and data sharing are attracting more attention from governments, funding bodies, academic and research institutions and researchers.

However, for HEIs especially in Africa to achieve the full benefits of data management they must go beyond developing data management policies to establishing appropriate infrastructure and services to support effective RDM (Jones, Pryor & Whyte, 2013:2). Valuable lessons can be gleaned from the various implementation strategies of international pacesetters, in particular, Australia, the UK and the USA (Henderson & Knott, 2015; Ball, 2013; Jones, Pryor & Whyte, 2013; Wilson et al., 2011; Takeda et al., 2010).

1.2. Institutional context: the University of Ghana, Legon

The University of Ghana (UG) is the premier and the largest public university in Ghana with multiple campuses from where it runs its academic and research programmes. Since its establishment in 1948, UG has undergone remarkable transformations in terms of its governance structure, institutional processes, human resource development, infrastructure development, academic programmes and institutional aspirations (UG, 2014; Daniel et al., 2007). With its current student population of nearly thirty-eight thousand, and faculty members of about one thousand two hundred, UG has evolved from merely a leading tertiary institution in Ghana to one of the leading universities in Africa (UG, 2016) with a reputation for being a centre for academic and research excellence (ORID, 2016).

In 2014, UG launched a new strategic plan which outlined nine strategic priorities which will provide direction for the University community and management team for the next decade to achieve its vision of a world-class research-intensive university (UG, 2014:9). The first of these priorities is to create an enabling ambience for high quality research and community engagement. Attracting more research funding (at least 150% increase) and quadrupling the number of faculty member's publications in high-impact journals as well as "[e]stablishing and joining global research networks" (UG, 2014:12) are some of the aspirations of the University for the next decade. The University in its commitment to this priority continues to implement initiatives and policies that would raise its research profile. The establishment of the Office of Research, Innovation and Development (ORID), headed by a Pro-VC and tasked

with facilitating, coordinating and raising funds to support research at UG (ORID, 2015) and the subsequent development of policies and systems to guide high quality research conduct and data management are indicative of this commitment.

At the beginning of the 2014/2015 academic year and in line with its strategic priority of overhauling and realigning institutional processes, UG changed its governance structure by adopting a collegiate system with all the academic units and some administrative functions reorganised along four colleges as follows: College of Health Sciences, College of Basic and Applied Sciences; College of Humanities and the College of Education (UG, 2016). This decentralisation of the governance and management structure not only paves the way for greater effectiveness and efficiency in its governance system but also, as stated by the then Vice-Chancellor Professor Ernest Aryeetey, "[it] has provided new opportunities for Schools and Institutes to carry out cutting-edge research with collaboration from local and international partners" (ORID, 2016:6). He adds, that each of the four Centres of Excellence had been allocated GH¢ 1 million (approx. $260,000) as seed money to facilitate research development in the areas of "Climate Change Adaptation, Malaria Research, Food Production and Processing, and Development Policy and Poverty Monitoring and Evaluation, to address key developmental issues in Ghana and beyond" (ORID, 2016:6).

Since 2014, ORID has been publishing an annual report of collective research and innovation efforts within the University called the *University of Ghana Research Report*. This report also highlights the numerous funding agencies that are supporting scholarship and innovation at UG in addition to internal funding opportunities available for research. Notable examples of external funders that work with the ORID include the NIH, NSF, Research Council UK (RCUK), and the Wellcome Trust (ORID, 2016:7).

UG has also invested in an elaborate network infrastructure that supports and promotes excellent teaching, learning and research, as well as administrative processes, with a current bandwidth capacity of 1 gigabits per second. This is managed by the University of Ghana Computing Systems (UGCS) which is also responsible for campus computing facilities and sites. The University of Ghana Library Systems (UGLS), which consist of a central library (Balme Library) and a number of college, school, departmental and residential libraries in the UG academic community, is at the heart of scholarship at UG. This unit has been supporting researchers and academics through the provision and promotion of information services and resources in a range of formats.

The UG recognises that developing and promoting cutting edge research is how the university can remain relevant in supporting policy making at the national level as well as contributing to the global developmental agenda (ORID, 2016:5; UG, 2014:9). Yet, this cannot be done sustainably without external funding support. UG has been more concerned with achieving high quality research outputs (findings) and their application to solving practical problems (technology transfer), than on the management of research data, though

it expects researchers doing research under the auspices of the University to adhere to high ethical and data management standards (UG, [2013]:3,8).

However, as global trends indicate that funding partners in the near future will tend to become keen on data management commitment, and the fact that effectively managing institutional data assets has far-reaching benefits, the emerging global and data intensive outlook of research and the ambition of the University to join the ranks of world-class research-intensive universities provide sufficient motivation to take initiatives in this emerging area of eScience and to develop an institutional data management framework that will consolidate all the infrastructure, systems and services at UG to support the divergent data management requirements that may exist across the University. Achieving this must start with an understanding of what is the current situation with regards to institutional practices, capacity, capabilities and preparedness to support RDM as well as the overall level of development in order to focus efforts in addressing these opportunities and challenges.

1.3. Problem Statement

RDM as a field of scientific enquiry is still evolving with much of the discussions on the subject being concentrated in the developed world. The extant literature on RDM development at HEIs have typically been projects that reported on the approaches and strategies adopted by institutions to implement RDM (Henderson & Knott, 2015; Hiom et al., 2015).

Henderson and Knott (2015) reported on the step by step approach adopted by the Virginia Commonwealth University Library (VCUL) for developing a data management programme. This included an assessment of institutional resources and existing services, identifying gaps in those resources and services and incorporating that into future RDM development. The authors, however, provided no details on other aspects of RDM development such as governance and resource availability. The study concluded that there are useful lessons to glean from the RDM community which can help to eliminate complications and build on the work that has already been done in RDM development. A similar study by Hiom et al. (2015) examined the approach used by the University of Bristol Library in implementing an RDM service. The study outcomes included an outline of the range of services developed in the course of the project. The authors concluded that, Universities must commit to investing in all aspect of RDM development including technical infrastructure, policy framework and human resources so as to take advantage of the opportunities for funding and partnerships (Hiom et al., 2015:491).

Whilst these studies have given an insight into how institutions have implemented RDM so far, studies that holistically assess the level of RDM development or the readiness of an institution to develop RDM are scares in the literature. In this study, the researcher looks at an institution that has not formally develop an RDM programme, to assess at the current

state of affairs in terms of RDM activities, capabilities, support and institutional preparedness at UG.

1.4. Aim and Objectives

1.4.1. Aim

The overarching aim of the study is to assess the state of the art of RDM development and institutional preparedness at UG in order to develop a roadmap for implementation at UG.

1.4.2. Objectives

1. To assess data management requirements, activities and capabilities at UG.
2. To benchmark existing practices, systems and facilities against that of the Engineering and Physical Sciences Research Council (EPSRC) policy framework to identify any gap in terms of institutional awareness, support, capacity and capability.
3. To identify priority areas for RDM development based on identified strengths and weaknesses.
4. To propose an action plan for RDM implementation at UG.

1.5. Research questions

The following questions will help identify the current state of RDM development and institutional readiness to support RDM at the UG:

1. Are there existing, well communicated, and well implemented policies, procedures and systems to ensure effective RDM at UG?
2. Are there adequate technologies (hardware and software) and security infrastructure to support effective RDM at UG?
3. Are there sufficient resources (finance, human, skills and competence) to support effective RDM at UG?

1.6. Scope, Limitations and Assumptions

The study focused specifically on RDM at the UG; because governance and legal frameworks, vision, culture and technical capabilities fundamentally shape how RDM is practised or supported within different institutional contexts, the findings of this study cannot be generalised to every HEI unless there is considerable contextual proximity with UG. Participants in the study were limited to the Library, Research Office, IT Services and faculty members in addition to a collection of institutional documents and policies relating to research and data management. These have been identified in the literature as major players in defining, developing and implementing RDM at the institutional level. The University management were not included mainly because their role is more at the strategic level, which involves, for instance, ratifying policies, approving plans and committing to RDM budget. This study was to assess data management requirements, activities and

capabilities and such information can best be elicited from data originators (researchers) and service providers. Ultimately the goal is to make recommendations that can be considered by management.

Institutional RDM requirement assessments often involve gathering requirements across different research and disciplinary groups. However, due to time and resource constraints, it was not feasible to conduct an institution-wide assessment of such groups. However, giving that this was a relatively small-scale study, two senior researchers were included, to get some input from the perspective of data originators and also as a way of initiating a consciousness and introducing faculty members to the idea of RDM.

1.7. Justification for the study

Onyancha (2016) in a study to investigate what is the current situation in sub-Saharan Africa in terms of research data publishing compared to international practice, found that sub-Saharan Africa accounted for 0.03% of the world's shared research data with more than 60% of that data coming from South Africa alone. This is no surprise as the literature reveals a dearth of RDM programmes and interventions on the African continent particularly at the institutional levels, which is the major contributor to open research data according to the findings of Onyancha (2016:10). What is more, there is little mention of RDM in the literature on Ghana. The remarkably few studies that have mentioned Ghana have either focused on the much broader concept of open data – as in "the Ghana Open Data Initiative [GODI]" (Ohemeng & Ofosu-Adarkwa, 2015) – and/or on the continental level (focusing on Africa in general rather than on Ghana specifically) such as the study by Onyancha (2016). The purpose of this study is in part, to fill this gap in the literature by contributing to the African story on RDM development in HEIs. van Deventer and Pienaar (2015:43) noted that initiatives and studies from developing countries can sometimes bring out insightful perspectives and revelations for experienced colleagues from developed countries.

The drivers for RDM development have been well documented in the literature and the responsibility to develop research data services, infrastructure and systems has been most emphasised at the institutional level (Awre et al., 2015:358; Wilson & Jeffreys, 2013:241; Jones, 2012a:49). Furthermore, the current global research landscape provides new opportunities and challenges for HEIs including effective and sustainable RDM. Global trends point to the growing interest of research funders in RDM: the importance of research data as an institutional asset is gaining traction among several top-notch HEIs, and the emerging global and data intensive outlook of research requires effective and secured systems to support such enterprise especially across distributed institutional research teams. The UG is not exempt from these issues, particularly as its ambition for the next decade is to join the ranks of world-class research-intensive universities. This provides sufficient motivation to develop an institutional data management framework that will consolidate all the infrastructure, systems and services at UG to support the divergent data management requirements that may exist across the University. In light of this, the research

topic was chosen because it enables the researcher to assess the current RDM activities, capabilities, support and institutional preparedness at UG. Benchmarking the findings against best practices helps to identify gaps based on which recommendations are made to the University management for consideration to implement an effective and coordinated data support service at UG.

1.8. Value of the study

The proposed study is significant and a timely one for UG in its renewed commitment towards becoming a world-class research-intensive university by 2024. The findings from this study may inform the University management and relevant stakeholders on the RDM capability, support and preparedness at UG. This will be useful particularly in the area of institutional planning and strategy development. So, also, is the benchmarking exercise which provides the opportunity to focus RDM development efforts at UG.

In view of current advances in and studies on RDM development elsewhere, the researcher believes that it is imperative to conduct this study to gain an idea of similar developments in Ghana's largest and most esteemed public university, the University of Ghana, Legon. The study, therefore:

1. contributes to the body of knowledge on institutional RDM implementation from the perspective of a developing nation;
2. contributes to the literature on the level of RDM development in Africa, specifically in Ghana;
3. fills a gap in the literature, as there is no recent research on RDM interventions and programmes on Ghanaian HEIs;
4. provides a useful case study and stimulates ideas for other institutions in Ghana who may want to implement RDM and participate fully in the global research agenda.

1.9. Terminology

1.9.1. Research data

What constitutes data is often not easily determinable and may connote different things across different domains and disciplines. However, for the purposes of this study, the definition of EPSRC will be adopted for research data. The council defines research data as "recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created" (EPSRC, 2015).

1.9.2. Research data management

Managing research data is an ongoing process throughout the lifecycle of the data. A widely cited definition of RDM in the LIS literature is proffered by the UK's Digital Curation Centre

(DCC) which defines RDM as "the active management and appraisal of data over the lifecycle of scholarly and scientific interest" (Jones, Guy, and Pickton, 2013: 5). Jones, Pryor and Whyte (2012:142) intimated that RDM essentially relates to "verifiable and replicable processes to support research data use from its planning, through its creation and active use, to its point of handover to a repository or archive, [but includes] preservation actions to ensure fitness for access, use and reuse". Also, Whyte and Tedds (2011:1) define RDM as "the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results [and] aims to ensure reliable verification of results, and permits new and innovative research built on existing information". Eckard (2014:6) also describes RDM as a data lifecycle process involving activities such as the data creation, data processing, data analysis, data preservation, data sharing and data reuse. Lastly, Pienaar (2011:8-9) asserted that there are many different definitions and terminologies associated with RDM, however RDM goes beyond backing-up data or archiving it but includes a variety of processes from "data ownership, data collection, data storage, data protection, data retention, data analysis, data sharing, and data reporting". In this study all activities that capture, document, organise, curate, preserve and provide continued access to research data during and after the research process constitute RDM.

1.9.3. Metadata

Often described as "data about data" (NISO, 2004: 1), the term metadata in this study means a systematic description of datasets which facilitates their sharing, discovery, re-use and citation.

1.9.4. Support and administrative services units

According to Jones, Pryor and Whyte (2013:3) the support and administrative units constitute the team that will actually deliver RDM services and includes, but is not limited to, the library, IT, records management and research administration functions. Each of these units has a role in the development and delivery of data services. In this study support and administrative services units or simply support services implies the Office of Research, Innovation and Development (ORID), the University of Ghana Library System (UGLS) and the University of Ghana Computing System (UGCS).

1.9.5. RDM services/Data service

In this study data services or RDM services means any services provided by support and administrative services units at any stage of the research cycle with the aim of supporting and promoting effective management and sharing of research data.

1.9.6. Data curation

According to Cragin et al. (2007) data curation is "the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education".

It involves inter alia describing, authenticating, representing and annotating research data through the application of appropriate metadata standard.

1.9.7. Digital curation

The UK's DCC defines digital curation as "maintaining, preserving and adding value to digital research data throughout its lifecycle [and involves] the active management of research data [to reduce] threats to their long-term research value and [to] mitigates the risk of digital obsolescence" (DCC, n.d.b). The activities involved in managing research data as well as curating it for long term preservation and discovery is collectively digital curation (Higgins, 2012:17). These activities are frequently done by a data steward (curator) whose role is to add value to the data, make it easier to discover, access, comprehend and reuse both for the short and long term (Carlson, 2014:65, Pennock, 2006:1).

1.9.8. Digital preservation

Digital preservation also refers to "the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value" (Pennock, 2006:1). It essentially involves the curation and archiving of research data (Berman et al., 2010:3). Whyte and Tedds (2011:1) asserted that whilst preservation ensures that data stored within repositories and at data centres remains accessible and usable in the longer term, curation ensures that data can be discovered, accessed and repurposed after its original use, both in the near and far future.

1.10. Organisation of the dissertation

The research report is organised into five chapters.

The first chapter presents the background to the study, a description of the institutional context, the research questions, aims and objectives, scope and limitations of the study, justification for the study, the value of the study and clarification of key terms.

The second chapter provides a review of the growing literature on RDM as a concept, the benefits and drivers of RDM, the global RDM landscape RDM in Africa, challenges and opportunities for HEIs, implementation strategies and the role of the library.

Chapter three deals with the methodology that was applied to this study. It provides a description of how the research was conducted, the approach, the strategy, the sampling methods, data collection instruments and procedures, data analysis method, trustworthiness of the research and ethical considerations.

In chapter four data is analysed and the findings are discussed. Chapter five provides a summary, recommendations and conclusion of the study.

1.11. Conclusion

In this chapter, a general background to the study was presented including a thick description of the University of Ghana which is the institutional setting for this investigation. The chapter also presented the research questions, aim and objectives of the study, the scope and limitations of the research work, justification and the value of the study, clarification of key terms, organisation of the research report and the time schedule that guided the conduct of this study. In chapter two, the emerging literature on RDM will be re-examined to unearth the meaning of the concept, its implications for HEIs and strategies for development and implementations.

CHAPTER TWO: LITERATURE REVIEW

2.1. Introduction

This chapter explores the growing literature on the subject of Research Data Management (RDM). The existing literature on RDM development at HEIs have mostly concentrated on developed countries and often provide insight into how institutions have implemented RDM so far, however, studies that holistically assess the level of RDM development and the readiness of an institution to develop RDM, particularly from a developing country are scares. Therefore, the literature review focuses on making a case for RDM development, as well as highlights pertinent issues bordering on implementation. The reviews expounds the concept of RDM and discusses several topical issues including benefits and drivers of RDM, the global RDM landscape, RDM in Africa, challenges and opportunities for HEIs, implementation strategies, role and responsibilities and competence requirements for the library, infrastructure and support services for RDM, and the EPSRC policy framework.

In order to address the aforementioned topics, relevant literature was retrieved from the Library and Information Science journals on the Emerald platform, Academic Search Complete (EBSCOhost), Library, Information Science and Technology Abstracts (EBSCOhost), Library and Information Science Sources (EBSCOhost), Library and Information Science Abstracts (ProQuest) and MasterFile premier (EBSCOhost) databases. Other relevant resources were retrieved from reliable websites through Google Scholar searches. E-books and hardbound books were also consulted. Searches were conducted from October 2015 through March 2016 and some of the keywords used include “research data management”, “data management”, “research data”, “data curation”. Materials selected for the review ranged from 2006 to 2016 and include peer-reviewed empirical studies and anecdotal articles (from authoritative sources such as the Digital Curation Centre [DCC]), government policy documents and commissioned reports. Materials were mainly on the experiences from Australia, South Africa, UK (predominantly) and the USA and only materials published in the English language were selected for the review. Materials were considered relevant for the review if they explain the RDM concept and provide justification for RDM, if they report on the role of the library, or if they report relevant or unique cases of RDM implementation programmes in HEIs.

2.2. Research data management (RDM)

RDM as a field of scientific inquiry is still evolving. As a domain within the emerging research paradigm of eResearch or eScience, RDM is increasingly attracting more attention from governments, funding bodies, academic and research institutions and researchers (Kahn et al., 2014:296; Qin, 2013:214; Pryor, 2012; Whyte & Tedds, 2011) for its value and prospects for advancing scientific research. The emerging research culture is a result of the growing influence of the digital revolution and is “characterized by data-intensive and networked collaboration” (Cox & Pinfield, 2014:299; Wang, 2013:1). Indeed, an attested feature of the

21st century research landscape is the proliferation of research data, growing exponentially and in many cases digitally-born (Ray, 2014:1; Charbonneau, 2013:365; Pryor, 2012:1, 7). The report of the "Blue Ribbon Task Force on Sustainable Digital Preservation" also acknowledged the surge in data intensive research and the need for sustainable research data:

> *There is a remarkable growth of data-intensive research in all knowledge domains. In most fields, there is high recognition of the benefits of preserving research data for various purposes and lengths of time. […] Research and education institutions, professional societies, archives, researchers, and the funding agencies that support data creation all have leading roles to play in creating sustainable preservation strategies* (Berman et al., 2010:3).

The proliferation is not only a result of the original data generated from carrying out a research activity but also reflects the possibilities of sharing that data which can be reused or mined to generate new data and findings (Kahn et al., 2014:296; Pryor, 2012:1). Wong (2009:125) stated that the recognition of this "data deluge phenomenon", the possibility of repurposing and reusing research data and the aspiration of maximising the return on public investments have all contributed to the surge in the discussion on RDM.

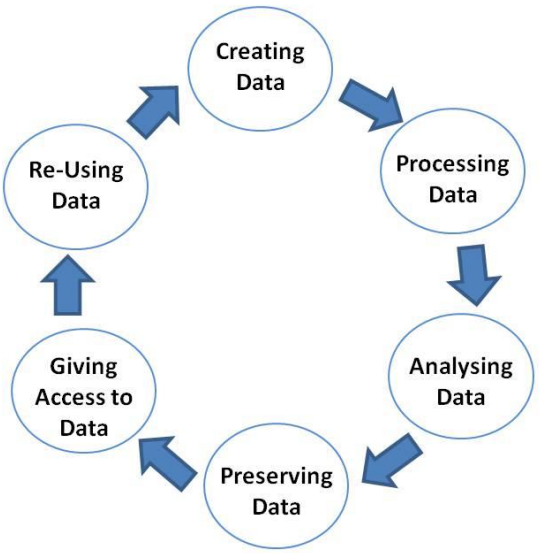
There seems to be a lack of clarity in the literature on the definition of RDM as a concept and terminology, partly as a reflection of the diversity in practice and the scale of institutional infrastructure and support. The term has been linked with other concepts like data curation, digital curation and digital preservation (van Deventer & Pienaar, 2015; Wang, 2013; Harris-Pierce & Liu, 2012; Whyte & Tedds, 2011), which have been used interchangeably in the RDM literature, thus further compounding this definitional conundrum. However, the relationship of RDM with the data and research lifecycles (Carlson, 2014; Wiggins et al., 2013; Higgins, 2012; Pryor, 2012; Strasser et al., 2012), and the need to manage research data (Holdren, 2013; Jones, Pryor & Whyte, 2013; Borgman, 2012; Harris-Pierce & Liu, 2012; Pryor, 2012; Strasser et al., 2012; Whyte & Tedds, 2011) have been widely acknowledged. Surkis and Read (2015:154) noted that the data lifecycle concept has been instrumental in explaining and understanding the "scope and meaning" of RDM and Pryor (2012:6-7) elucidated the centricity of data to the research lifecycle and the need for specialist skills to manage it.

To begin with, the scope of RDM definition varies. Some authors view RDM as a comprehensive concept that relates to "all aspects of the data lifecycle" including the appraisal, curation and archiving of research data for subsequent and long term preservation, discovery and re-use. This view held by authors such as Eckard (2014), Jones, Guy and Pickton (2013), Pienaar (2011) and Whyte and Tedds (2011) is corroborated by Charbonneau (2013:366) who writes "Data management is a comprehensive approach to effectively sharing, managing, curating, preserving, and reusing the rapidly growing volumes of data generated by research". In simple terms, RDM involves data related activities

performed during and after a research programme to make it fit for subsequent and long term reuse.

In contrast, other literature attempts to delimit the frontiers of RDM, and separates the activities that are carried out to support RDM during the active phase of the research activity from other data related activities, support and services offered after the research process as espoused by authors like Jones, Pryor and Whyte (2012). Carlson (2014:64-65) supports this perspective when he delineates data management and data curation services by emphasising a fundamental difference where the former relates to support given to researchers during the data-active phase only -- such as writing data management plans and providing training -- and the latter covering support for both the active phase of the data in a research programme and when the programme is completed, such as describing the data with appropriate metadata and assigning a digital object identifier (DOI) for continual discovery and contextualization.

It is also interesting to observe that, while Jones, Pryor and Whyte (2012:142) sought to distinguish RDM from preservation, their definition actually acknowledges preservation activities as being present; but preservation includes curation and archiving (Berman et al., 2010:3), which Carlson (2014) separates from pure data management activities. Again, Carlson's (2014:65) description of data curation as an ongoing management of data, as long as it is useful to the scholarly community, resonates with the comprehensive perspective of RDM held by Eckard (2014), Charbonneau (2013), Pienaar (2011) and Whyte and Tedds (2011).



Creating data
- design research
- plan data management (formats, storage etc)
- plan consent for sharing
- locate existing data
- collect data (experiment, observe, measure, simulate)
- capture and create metadata

Processing data
- enter data, digitise, transcribe, translate
- check, validate, clean data
- anonymise data where necessary
- describe data
- manage and store data

Analysing data
- interpret data
- derive data
- produce research outputs
- author publications
- prepare data for preservation

Preserving data
- migrate data to best format
- migrate data to suitable medium
- back-up and store data
- create metadata and documentation
- archive data

Giving access to data
- distribute data
- share data
- control access
- establish copyright
- promote data

Re-using data
- follow-up research
- new research
- undertake research reviews
- scrutinise findings
- teach and learn

Figure 2.1 Research Data Lifecycle (UK Data Archive, n.d.)

What is more, the use of the lifecycle model in describing RDM depicts a sequential flow of iterative processes that are connected and supported by dedicated resources, infrastructure and services. Carlson (2014) discussed how the lifecycle model serves as a framework for RDM particularly in cascading the relationships between the distinct stages of research data as well as developing support services. Donnelly (2012:101) also emphasised this point stating that "data management is a flow, a chain of events with multiple actors representing a variety of stakeholder groups". The RDM process encapsulates a multiplicity of activities relating to the capture, documentation, organisation, curation, storage, preservation and provision of continued access to data during and after the active research phase. These activities are best illustrated with a data lifecycle such as those used by Jones, Guy & Pickton (2013:5), Wiggins et al. (2013:2) and Strasser et al. (2012:3). Figure 2.1 shows an example of such a data lifecycle. The DCC Digital Curation Lifecycle model (Higgins, 2008) provides a generic framework which focuses on necessary activities and processes for curating research data throughout its lifecycle. Cox and Pinfield (2014:300), acknowledging these activities add that contextual factors such as "technical capabilities, ethical considerations, legal issues and governance frameworks" fundamentally shape RDM practices and support within different institutional contexts.

The literature also reveals that the way researchers manage and share their research data varies across disciplines (Tam, Fry & Probets, 2014: 722). Two landmark studies into researchers' data related practices and behaviours across selected disciplines, "*To share or not to share: publication and quality assurance of research data outputs"* (RIN, 2008) and "*DCC SCARP: disciplinary approaches to sharing, curation, reuse and preservation"* (Lyon et al., 2010) lend credence to this assertion. For instance, the Research Information Network (RIN) report shows that:

> *[t]here are significant variations […] in researchers' attitudes, behaviours and needs, in the available infrastructure, and in the nature and effect of [RDM] policy initiatives, in different disciplines and subject areas* (RIN, 2008:7)

and the DCC SCARP project stated:

> *[a] central finding is that institutional repositories responsible for curating data produced by their own research community will need to develop domain-specific strategies since a generic approach to data curation will not be sufficient to cope with the different data-related needs and expectations of researchers working in different disciplines* (Lyon et al., 2010:10).

Another study co-sponsored by the RIN and the British Library to "improve understanding of information use in the life sciences" (Williams et al., 2009:48) recognised a fundamental challenge with RDM and its policy-related implementations arising from confusion over

terminology. The report highlighted the fact that data itself can connote different things across different domains and disciplines. A consequence of such confusion is the often-encountered "conflation of information exchange with the sharing of research data" (Williams et al., 2009:49). Such inconsistencies across disciplines, for instance, have also contributed to the ambiguity of RDM as a concept and terminology.

Higgins (2012:22) outlined three data management approaches that may be visible in an institution:

1. The management and stewardship of research data rest with the research team during and after the research;
2. The management and stewardship of research data is undertaken by the research team, with support from information professionals such as librarians particularly in the area of advice on best practices, training, and data management planning; and
3. Data management is undertaken by the research team during the research-active phase after which it is handed over to another party (such as the organisation's library, a domain-specific library or a data centre) for stewardship and continuous management.

Whatever method is followed, the approaches identified by Higgins suggest that the management of research data does not only take place while the research is in progress but continues even after it is completed (where data is curated and archived); an assertion that follows in the line of authors such as Eckard (2014), Charbonneau (2013), Pienaar (2011) and Whyte and Tedds (2011) who consider RDM as a comprehensive process covering data management planning, data curation, preservation, sharing and the reuse of research data. More importantly, the benefits of managing and sharing research data has been widely acknowledged and professed at the national and international level (Holdren, 2013; OECD, 2007), funder level (Friedlander & Adler, 2006) and in scholarly discourse (Jones, Pryor & Whyte, 2013; Borgman, 2012; Whyte & Tedds, 2011).

2.2.1. Benefits and drivers for RDM

Justifications for good data management practices have been copiously documented in the extant literature. Studies discussing the topic have often cited a number of drivers and benefits for managing and sharing research data. These drivers and benefits range from the data deluge phenomenon and its concomitant challenges as well as prospects for advancing scientific research; disciplinary protocols; national and international policies; directives of funding bodies; and the economic implications for wringing out the optimum value from research data. The OECD principles and standards for accessing publicly funded research data recognise that good data management and sharing are requisites for;

1. *good stewardship of the public investment [into research];*

2. *creation of strong value chains of innovation [through repurposing and reuse of data]; and*
3. *enhancement of value from international co-operation [through collaborative research] (OECD, 2007: 9-10).*

Meanwhile, Pryor (2012:2-3) and Wong (2009:125) have taught us to reject any notion that the need for data management as an integral part of the scientific research process and workflow can only have pertinence in the science and technology domains, reinforcing the universality of the need to manage research data.

2.2.1.1. Benefits

Managing and sharing research data have benefits for the researcher, the broader scientific community, academic and research institutions and the public at large (Holdren, 2013:1; Harris-Pierce & Liu, 2012:599; Strasser et al., 2012:1-2).

For researchers, good data management enables them to plan how they will gather, document, process, organise, manage and preserve data before the start of the actual research project. This saves them time and energy which can be focused on the main research work. This also has implications for conducting high-quality research (Jones, Pryor & Whyte, 2013: 2; Strasser et al., 2012:1). The quality and integrity of research is an issue of keen interest for policy makers (Whyte & Tedds, 2011:2), especially because government agencies are under pressure to demonstrate value for investing public monies into such enterprise.

In collaborative research environments, good data management enables other collaborators in a research team to access, understand and use data (Strasser et al., 2012:1). Jones, Pryor and Whyte (2013:2) added that beyond the research-active phase, good data management makes data discoverable, accessible and intelligible for reuse in the long term. Strasser et al. (2012:2) agree, avowing that the comprehensive description of datasets and the contextual details provided can be useful for future researchers to understand the data and use it. What is more, when data is well managed, it can be shared, thereby increasing the visibility of researchers which impacts their bibliometric and altmetrics indices (Jones, Pryor & Whyte, 2013:2; Strasser et al., 2012:1).

For the broader research community, Borgman (2012:1070) argues that well managed and shared data encourages meta-analysis (merging data from different sources and time to draw new hypotheses). This also implies that new inferences can be drawn from data originally collected for other purposes. Harris-Pierce and Liu (2012:599) corroborate this assertion stating that well managed data could become a vital part of the science ("through data-mining and reuse", and data visualisation) and the sharing could "lead to discovery and use outside the discipline in which the data was created, fostering interdisciplinary research

and learning". The OECD report further emphasises this point as well as other specific benefits for improved access to and sharing of research data (OECD, 2007:9-10).

At the institutional level, managing research data can help mitigate any risk to the reputation of an institution (Whyte & Tedds, 2011:2) because good data management ensures that researchers adhere to high standards of research. Again, the access and use of well-managed research data (which is an institutional asset) saves institutions from spending extra money to generate new data (Harris-Pierce & Liu, 2012:599).

Open access to research data has benefits for the general public who in reality are the true funders of such enterprise. The Office of Science and Technology Policy (OSTP) in a memo to Heads of Executive Departments and Agencies in the USA noted that:

> *[a]ccess to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather data underpins the forecasting industry, and making genome sequences publicly available has spawned many biotechnology innovations* (Holdren, 2013:1).

Yet open access per se does not guarantee quality and usability of data: good data management practice ensures that what is eventually shared is comprehensible and fit for re-use.

2.2.1.2. Drivers

According to Whyte and Tedds (2011:2) motivation for data management and sharing hinges on the principle that publicly funded research and, for that matter, the data generated from it is public good. In the UK for instance, an estimated £3.5 billion public money is invested annually as funding for research carried out in UK's HEIs (Pryor, 2012:1). As a result, this principle has been upheld by policy makers and reflected in amendments to legislation to support this course of action. A number of factors contributing to this drive particularly at the national and institutional levels are discussed below.

2.2.1.2.1. Global and data intensive outlook of research

Jones, Pryor and Whyte (2013:2) argue that the current research landscape is characterised by networked collaborations supported by a growing cyber-/e-infrastructure where researchers and their collaborators must share and exchange data over various platforms within the virtual research environment (VRE). In the UK, in particular, funders are encouraging researchers to collaborate more on the international level (Whyte & Tedds, 2011:2). This will require effective and secured systems to store, access, and share data and other tools with colleagues across different institutions and locations. Good data management is also necessary to ensure that collaborators can understand what data is shared.

2.2.1.2.2. Research quality and integrity

Whyte and Tedds (2011:2) stated that policy makers are keen on the quality and integrity of research. Publicly funded research mostly aims to shape national policy and developmental programmes and the veracity of the study must underscore the viability or otherwise of its output (findings). For institutions where research is carried out, such underpinnings have an impact on their reputation. Lawmakers in the UK, for instance, mandate institutions to ensure that research conducted by their staff conform to the highest standards including the gathering, documentation, processing, management, storage and sharing of research data (Whyte & Tedds, 2011:2). This makes it possible to reproduce or verify a scientific study (Ray, 2014:7). The reproducibility or verifiability of research is not only useful for testing or validating the science but also confirms that taxpayers' monies are well spent. (Borgman, 2012:1067).

2.2.1.2.3. Research funders' data policies

Increasingly, funding bodies such as the Research Council UK (RCUK), National Science Foundation (NSF), USA, National Institutes of Health (NIH), USA and Australian Research Council (ARC) are making data management and sharing a key element in the selection criteria for awarding research funding (Borgman, 2012:1060; Jones, 2012a:48-89; Whyte & Tedds, 2011:2). The DCC provides a summary of funder policies in the UK (Jones, 2012a:50-53; Jones, 2012b) and the prominent themes are development of data and documentation; legal and/ or ethical strictures; data sharing; and preservation and long-term accessibility (Jones, 2012a:54).

The EPSRC in the UK mandates institutions to develop policies and infrastructure for RDM support, and data from funded research must be curated and preserved for not less than ten years. Data management plans are also a prerequisite for accessing funding (Whyte & Tedds, 2011:2). In the USA, the NSF since 2010 requires a two-page data management plan (to be peer reviewed) to be added to grant proposals and the NIH, since 2003, requires data management plans for grant requests above $500,000 (Borgman, 2012:1060). Aside from the *Australian Code for the Responsible Conduct of Research* (Australian Government, 2007:2.1) policy document (jointly issued by NHMRC, ARC, and universities in Australia), which promotes proper management and retention of research data, the Australian National Health and Medical Research Council (NHMRC) Statement on Data Sharing further states:

> *NHMRC-funded researchers are encouraged to consider from the earliest stages of research planning how their research data will be obtained, managed, curated, stored and disseminated into an appropriate, publicly accessible database* (NHMRC, 2016).

2.2.1.2.4. Journal policies for research data deposit and access

Another emerging trend in the modern research landscape has to do with some journal publishers' requirement for "data and other research documentation associated with published articles" to be published or deposited (Borgman, 2012:1060; Strasser et al., 2012:2). This may be subject-specific data repository such as *Protein Data Bank* (http://www.rcsb.org/pdb/home/home.do) or into public archives. Table 2.1 shows a list of some journals that have data deposit requirements.

Name of Journal	Data deposit requirement
Nature	http://www.nature.com/authors/policies/availability.html
Public Library of Science (PLOS)	http://journals.plos.org/plosone/s/data-availability
Science	http://www.sciencemag.org/authors/science-editorial-policies
The American Naturalist	http://www.journals.uchicago.edu/journals/an/instruct#data
Evolution, Journal of Evolutionary Biology	http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291420-9101/homepage/ForAuthors.html

Table 2.1 List of some journals that have data deposit requirements

2.2.2. The global RDM landscape

The subject of RDM and data sharing has gained traction particularly among OECD countries. These countries have in most cases enacted legislation and/or issued policy directives to enforce the proper management, preservation, sharing and open access to research outputs including research data that are state-funded. The compliance of such directives as the OSTP memo to Heads of Executive Departments and Agencies (Holdren, 2013) has led to funder policies and requirements for RDM. Jones (2012a:58-59) noted that while varying degrees of such compliance is promulgated through different funder policies in countries such as the UK and USA, others like Australia have adopted a collaborative national approach in which common regulations and requirements are outlined to guide individual (researcher) and institutional RDM practice. This section provides a brief picture of the global RDM landscape, focusing on three OECD member countries (Australia, United Kingdom, and United States of America). Other notable countries include Canada, Finland, Germany, Netherland and New Zealand (Jones, 2012a:61).

2.2.2.1. Australia

In 2003, the Australian Government appointed a working group made up of representatives from the Australian Research Council (ARC), National Health and Medical Research Council (NHMRC) and the Australian Vice-Chancellors' Committee (AVCC). The committee was tasked with reviewing the existing Joint NHMRC/AVCC Statement and Guidelines on Research Practice (1997). In 2007 the committee presented its report titled, *Australian Code*

for the Responsible Conduct of Research [currently under review]. The report sets out institutional and researcher responsibilities and guidelines for managing research data. These include providing a secure infrastructure to support RDM as in data storage and record keeping, establishing policies to guide data management, data sharing and retention, and data ownership (Australian Government, 2007:2.1-2.3). Currently, compliance with the code is a condition for accessing ARC and NHMRC funding, and is expected to extend to other government and funding institutions soon.

Many universities in Australia have responded to this code by putting together an institutional RDM policy and initiating some support service through their libraries. Jones (2012a:59) highlighted some of these institutions. Table 2.2 shows some of these institutions and links to their RDM policies and Library support web page.

Institution	RDM Policy	Library support for RDM
Queensland University of Technology [QUT]	http://www.mopp.qut.edu.au/D/D_02_08.jsp#D_02_08.01.mdoc	https://www.library.qut.edu.au/research/data/
University of Melbourne	https://policy.unimelb.edu.au/MPF1242	http://researchdata.unimelb.edu.au/help
Monash University	http://policy.monash.edu.au/policy-bank/academic/research/research-data-management-policy.html	https://www.monash.edu/library/researchdata

Table 2.2 Australian Universities RDM policy and Library support

In addition, the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) programme is investing into the Australian National Data Service [ANDS] (http://www.ands.org.au/) which is transforming Australia's research data environment through a national research data commons, and adding value to data assets by effectively managing them and providing tools and capabilities for aggregating, connecting, enabling discovery and supporting their reuse (Treloar, Choudhury & Michener, 2012:173-192; ANDS, n.d.).

2.2.2.2. UK

According to Jones (2012a:58) a £4.3 million Joint Information Systems Committee Managing Research Data (JISCMRD) programme (http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx) from October 1, 2009 to September 30, 2011 has been influential in the advancement of RDM strategies in the UK. This first programme addressed issues on infrastructure development; RDM planning; tools and support for RDM; citing, linking, integrating and publishing (CLIP) research data; and RDM training (Jones, 2012a:58; JISC, n.d.a). In collaboration with the Digital Curation Centre (DCC), a second programme

which ran from October 3, 2011 to July 31, 2013 focused on improving the capacity and capability of UK universities to support RDM through local level policies and the development of data services and systems (Jones, Pryor & Whyte, 2012:142; JISC, n.d.b).

Furthermore, the Research Council UK (RCUK) issued a seven point statement of common principles on data policy which provides a comprehensive framework for individual Research Council policies on RDM (Jones, 2012a:48; RCUK, 2011). In 2011 one such organisation, the Engineering and Physical Sciences Research Council (EPSRC) released its data policy spelling out nine expectations (last updated October 9, 2014) towards the management and access of EPSRC-funded research data (EPSRC, 2011), with May 1, 2015 as deadline for full compliance or lose eligibility for EPSRC funding. The policy places the responsibility on institutions to develop appropriate data policies, infrastructure and services and ensure that their researchers are aware of these as well as their individual responsibilities regarding the management and sharing of research data (Wilson & Jeffreys, 2013:241; Jones, 2012a:49). These expectations have also been used in many variants by institutions in the UK to scope and benchmark existing RDM services and capabilities. The DCC, working with HEIs within the UK, has also developed several templates and transferable toolkits to support RDM.

2.2.2.3. USA

According to Henderson and Knott (2015: 48) the "National Science Foundation (NSF) Digital Libraries Initiative (1994–2003)", "National Institutes of Health (NIH) Data Sharing Policy (2003)" and the "Association of Research Libraries (ARL) report to the NSF, *To Stand the Test of Time: Long Term Stewardship of Digital Data Sets in Science and Engineering* (2006)" provided the foundation for digital and data initiatives in the USA. In February 2013, OSTP Director, John Holdren, issued a directive to heads of executive departments and agencies mandating requirement for "data management plans and access to research data for federally funded scientific research" (Holdren, 2013:5). As a result, major research funding bodies such as the NSF and NIH have begun requesting for data management and sharing plans as an addendum to the requirements for grant application (Jones, 2012a:61; Strasser et al., 2012:2). Bishoff and Johnston (2015) provide an analysis of NSF data management plans (DMPs) at the University of Minnesota. Jones (2012a:61) also outlines a number of institutional efforts towards supporting USA university researchers in meeting these mandates; for instance, the California Digital Library's compendium of funder policies and DMP requirements (https://dmptool.org/guidance), data management guidance support at MIT Libraries (http://libraries.mit.edu/data-management/) and the ARL Guide to librarians (http://old.arl.org/rtl/eresearch/escien/nsf/index.shtml) to better understand NSF data sharing policy and offer better support to researchers on data management planning.

Furthermore, two major projects, Data Observation Network for Earth [DataONE] (https://www.dataone.org/) and Data Conservancy [DC] (https://dataconservancy.org/),

both initiated and funded through the NSF Sustainable Digital Data Preservation and Access Network Partners (DataNet) programme since 2009, have developed data curation tools and services to support the interdisciplinary and networked research communities, a cyberinfrastructure to support these services, models for sustainable research data and also programmes for capacity building and professional development (Treloar, Choudhury & Michener, 2012:194-201).

2.2.3. RDM in Africa

The literature reveals a dearth of RDM programmes and interventions on the African continent particularly at the institutional levels. Except for South Africa's distinct leadership in higher education on the African continent, no other country was found making efforts towards this new domain of e-science. As testament to this leadership role, South Africa has made inroads in terms of recognition by the National Research Foundation (NRF) through its policy statement, and the development of national infrastructure such as the South African Data Archive (SADA) and the Data Intensive Research Infrastructure for South Africa (DIRISA) initiatives. There are also institutional efforts from University of Pretoria (UP), Council for Scientific and Industrial Research (CSIR) (van Deventer & Pienaar, 2015:35), Stellenbosch University, University of Cape Town (UCT), University of South Africa (UNISA), University of the Witwatersrand (Kahn et al., 2014: 297), the list continues, who have developed RDM related policies and codes on good research practices and ethics.

Several empirical and anecdotal studies have been conducted on the South African RDM experience (Denny et al., 2015; van Deventer & Pienaar, 2015, Kahn et al., 2014; Lötter, 2014; Pienaar, 2011). An account by van Deventer and Pienaar (2015) explains RDM development at UP and CSIR and the lessons learnt. Among the important lessons learnt they noted that while valuable lessons abound from early adopters and international pacesetters, institutions implementing RDM interventions must consider their local and cultural context, ensuring that it is appropriate and acceptable for their environment (van Deventer & Pienaar, 2015:43). This corroborates the view of Cox and Pinfield (2014:300) who asserted that "technical capabilities, ethical considerations, legal issues and governance frameworks" fundamentally shape how RDM is practised or supported within different institutional contexts. Furthermore, initiatives from developing countries can sometimes bring out insightful perspectives and revelations for experienced colleagues (van Deventer & Pienaar, 2015:43). They also avowed that actively cultivating a network of researchers from different disciplinary backgrounds, and involving libraries when developing a cyber-infrastructure plan is a probable best approach for beginners (van Deventer & Pienaar, 2015:36). The establishment of the Network of Data and Information Curation Community (NeDICC), which is a community of practice (COP), has also complemented efforts towards the promotion of digital curation and RDM among practitioners and has facilitated their skills development and capacity building (van Deventer & Pienaar, 2015:36, 42; Kahn et al., 2014: 304). In concluding their narrative, van Deventer and Pienaar (2015:44) stated that

RDM awareness is high, expertise has been developed, high level support and commitment has been gained (from policy makers) and infrastructure has been developed to support RDM in South Africa. What is left is for the South African research community, including academic and research libraries, to harness these opportunities to participate fully in the global research agenda.

Interestingly, this contrasts with the findings of Kahn et al. (2014:299-300) who surveying forty-one participants at the Library and Information Association of South Africa (LIASA) workshop held in collaboration with the DCC report that twenty of the participants rated as average their "understanding and confidence in supporting […] RDM" and as many as thirty-one expressed lack of awareness of any existing national level RDM policy and framework. Despite these revelations Kahn et al. (2014:300) insist that the levels of RDM awareness were higher than anticipated. Other thematic discussions based on the findings -- "policy, policy-making and responsibilities", "skills and resource gaps" and "data sharing and Open Data" -- revealed less desirable results and required further attention (Kahn et al., 2014:303-304). The study made recommendation for addressing the pitfalls identified. More importantly, they noted that, despite the non-existence of a national or funder mandate for RDM, the sheer willingness, proactiveness and enthusiasm for addressing the RDM challenge was apparent and worth emulating by HEIs across the globe (Kahn et al., 2014: 305).

Lötter (2014) reflecting on the state of the art of RDM in South Africa corroborated many of the findings reported by van Deventer and Pienaar (2015) and Kahn et al., (2014). For instance, the enthusiasm and interest of several key stakeholders such as HEIs, research councils, government agencies and libraries and the surge in RDM initiatives such as the development of a national cyber-infrastructure to support data intensive research (e.g. Data Intensive Research Initiative of South Africa [DIRISA], and South African National Research Network [SANReN]), COP (e.g. NeDICC), and international participation (e.g. NRF membership in the Committee on Data for Science and Technology [CODATA]) (Lötter, 2014:5-6). She also noted that while the awareness perception of RDM is high, perception of maturity on other aspects such as "policies" and "staff skills" is relatively low (Lötter, 2014:16).

Notwithstanding that the RDM programme and intervention in South Africa are less than fully matured, particularly when compared with other countries such as the UK, USA and Australia, these studies reveal that South Africa has made significant advancements in the RDM agenda. Indeed, awareness is generally high, there is high level national support and commitment in the form of investments into infrastructure to support data-intensive research. Furthermore, HEIs are developing policies on RDM and COPs on data and information curation activities are existent. Still more attention is needed in the area of policy (national level), skills and resource development. High level coordination on RDM leadership and policy development (Kahn et al., 2014:304), as well as collaboration among

stakeholders in the areas of building capacity, sharing expertise and experiences, and supporting common objectives (Lötter, 2014:19) are just some of the ways South Africa can push boundaries and cement its continental leadership in the RDM discourse.

2.2.4. RDM in Higher Educational Institutions (HEIs)

In section 2.2.1 a number of benefits and drivers that provide justification for RDM were discussed. Providing a general backdrop, good data management practice was seen to have immense value for key stakeholders (researchers and the research community, HEIs and the public at large). Particularly for HEIs however, the implications of RDM transcend the apparent returns. Two reports by Beagrie, Lavoie and Woollard (2010) and Beagrie, Chruszcz and Lavoie (2008) confirm many of the institutional benefits and illustrate the existence of cost variables as well. This presents new challenges and opportunities for HEIs especially in the face of uncertain financial conditions. Particularly apparent are the interdisciplinary conundrum and sustainability concerns that arise as HEIs deliberate on RDM investments and implementation strategies.

To begin with, the responsibility to develop research data services, infrastructure and systems has been most emphasised at the institutional level (Awre et al., 2015:358; Wilson & Jeffreys, 2013:241; Jones, 2012a:49) as evident through the EPSRC policy framework (EPSRC, 2011), even though one can point to some national level infrastructure investments (e.g. DIRISA, ANDS and UK Data Archive) and capacity building support (e.g. DCC). Awre et al., (2015:358) assert that though appreciable efforts have been made in the area of policy development (at the national, funder and institutional levels), translating it to the level of local support service has been considerably slower. They blame this on the intricacies of the context within which HEIs must develop data services; among which they mention “the diversity of research” (i.e. the varying disciplinary and sub-disciplinary research cultures and their perception of what constitute data), and the challenges of changing researchers’ everyday research practices and behaviours (Awre et al., 2015:358), especially their attitude towards data management and sharing.

To corroborate and further advance this point, Procter, Halfpenny and Voss (2012:139-140), reviewing a number of reports and case studies such as the MaDAM project at the University of Manchester on the disciplinary difference in data management practices and behaviour, opine that one important element that will drive a successful adoption of HEIs’ research data management policy is to ensure that it is aligned with researchers’ everyday practices for a seamless fit into the research data lifecycle as depicted by the DCC lifecycle model. What is more, while a cultural change is necessary in many cases, it cannot be simply achieved by national, institutional or funder policy mandates, however, incentivising researchers at all levels (group, institution and community) to encourage an open data sharing culture, or developing RDM systems to support existing informal sharing practices might provide a head-start to evolving this cultural change (Procter, Halfpenny & Voss,

2012:139-140). The following authors further shed light on the varied research practices of researchers and across disciplines and the implication for data management: Whitmire, Boock and Sutton (2015); Weller and Monroe-Gulick (2014); Akers and Doty (2013).

Furthermore, justifying RDM infrastructure and data services investments at HEIs will require taking into consideration the cost-benefit implications. The reports of two studies funded by the UK's JISC; "*Keeping research data safe 2 [KRDS2]"* (Beagrie, Lavoie & Woollard, 2010) and "*Keeping research data safe: a cost model and guidance for UK universities"* (Beagrie, Chruszcz & Lavoie, 2008) have not only illustrated these cost variables, but have included a costing framework and guidance for UK universities. One can suppose that the cost-benefit justification for RDM infrastructure and data services investments at this point is no longer pressing (Procter, Halfpenny & Voss, 2012: 145) as it would be for sustainability. According to Lavoie (2012:68) sustainability can be conceived from three perspectives;

1. Technical sustainability: involves the development of RDM infrastructure, tools and processes that are resilient and scalable;

2. Social sustainability: involves a "shared commitment" to RDM from stakeholders (researchers, HEIs etc.) who have a common interest in the management, curation, preservation, future discovery and re-use of research data; and

3. Economic sustainability: involves ensuring that there is a regularly flow of funding, and all necessary resources to support RDM activities are adequate and continually available.

Sustainability aims to "allow scientists to build confidence in the availability of core technologies, software, data and services encouraging adoption and widespread use" (BIS, 2010:16). The focus here will be on economic and technical sustainability. As we find in the UK for instance, many of the RDM infrastructure and services development as well as capacity building in HEIs had funding and support from the JISC and DCC at the initial stages. A few of such projects are the data.bris project at the University of Bristol (Hiom et al., 2015), the MaDAM and Storage, Archive and Curation (SAC), projects at the University of Manchester (Procter, Halfpenny & Voss, 2012) and the "Embedding Institutional Data Curation Services in Research (Eidcsr)" and "Supporting Data Management Infrastructure for the Humanities (Sudamih)" projects at the University of Oxford (Wilson et al., 2011). The concern however, is how projects such as these can continue in a self-sustaining way after the expiration of the funding that was used to commence them. Another BIS report in 2011 emphasised this point, stating that concerns over the "ongoing cost of data curation" are common among HEIs, as funder requirements for long-term preservation of research data for instance often exceed "significantly […] the grant or award that first generated them" (BIS, 2011:11).

The DCC's *2014 RDM Strategy to Action Survey* shows that most institutions surveyed (approximately 90%) depended on internal sources of funding for new RDM-related appointments, "training and development" and infrastructure support (Whyte, 2014). However, in the face of uncertain financial conditions, institutions are looking to develop infrastructure and services that are income generating or at least capable of recovering the cost (Wilson & Jeffreys, 2013:242; Wilson et al., 2011:279). A number of funding models have been explored in projects such as the ones listed earlier. For instance, funding the cost of providing RDM infrastructure and services through institutional budget, payment by researchers or payment by funders as part of research grants they dole out [cost recovery model] (Wilson & Jeffreys, 2013:242).

The most prominent of these models has been the Direct Cost Recovery (DCR) model, for recouping cost through research grants (from Research Councils) based on the full economic costing (fEC) methodology (Hiom et al., 2015:484; Wilson & Jeffreys, 2013:242; Procter, Halfpenny & Voss, 2012:146). To do this, researchers include as part of grant proposals, all costs (direct or indirect) arising from the research project and in the management of research data. This will include inter alia any cost for using RDM resources, infrastructure and services such as data curation or storage. In that case researchers will have to provide an ex ante estimate of volume of data to be generated and storage capacity required (Procter, Halfpenny & Voss, 2012:146). The advantage with the fEC method is that it provides a reliable and transparent basis for costing HEIs' research infrastructure and services (Procter, Halfpenny & Voss, 2012:146), however, it also presents a higher risk of under-quoting or over-quoting prices (Wilson & Jeffreys, 2013:245), leading to the possibility of Research Councils declining due to the wide variations in HEIs' indirect costs for instance (Procter, Halfpenny & Voss, 2012:146).

Another model that has been explored particularly in the MaDAM, SAC and Data Management Roll-out at Oxford [Damaro] (Wilson & Jeffreys, 2013:236, 242) projects is the "free at the point of use" (FATPOU). This model provides "free access" to RDM infrastructure and services to the users (at the point of use), and the cost is recovered through "block grants" or "quality-related research funding" from funding councils such as the Higher Education Funding Council for England (HEFCE) (Procter, Halfpenny & Voss, 2012:146). The challenge with this is that it may resuscitate arguments over dual funding (Procter, Halfpenny & Voss, 2012: 146), especially at a time when HEIs are already levying high indirect costs on funding councils (Wilson & Jeffreys, 2013:242). At the University of Oxford, direct support for RDM from the University's budget was also considered (Wilson & Jeffreys, 2013:242), but again in the face of economic uncertainties, these might not be sustainable. These notwithstanding, Procter, Halfpenny and Voss (2012:146) noted that many of the costs and benefits of RDM remain unquantifiable and with respect to sustainably funding RDM, researchers and funders response to the "different charging models" is yet to be tested. Moreover, making RDM facilities and services "freely available" may attract many more researchers to utilise them and by extension effectively managing

and preserving data for future re-use, but charging for accessing these facilities may also provide a consistent and secured flow of revenue for continued RDM support (Lavoie, 2012:81).

Adding to the sustainability discourse, the constantly changing technological landscape poses a major decision making challenge for HEIs in implementing RDM infrastructure that is appropriate and sustainable. For instance, choosing between an internally developed storage infrastructure as against subscribing to a cloud service (Procter, Halfpenny & Voss, 2012: 146). Wilson et al., (2011: 278-279) provide a thorough discussion of the matter, highlighting important arguments that may arise as institutions deliberate on choosing to build an entirely new RDM infrastructure in-house when similar infrastructure exist at the national level which may provide better "economies of scale and concentration of expertise". What is important is for the pros and cons of each option to be weighed in order to make an informed decision (Procter, Halfpenny & Voss, 2012: 146). Using national infrastructure for instance may not provide the level of bespoke support needed for specific research projects but as Collins (2012:168-169) argues, they "represent an important and valued asset within the research infrastructure" and conspicuously play a critical role in supporting researchers and institutional RDM endeavour.

2.3. RDM Implementation, Infrastructure and Services Development in HEIs

2.3.1. Implementing RDM in HEIs

Henderson and Knott (2015:49) reporting on the RDM implementation experience at the Virginia Commonwealth University Library (VCUL) noted that "[b]y learning from the successes and mistakes of data management pioneers, some missteps could be avoided, some steps could be condensed or eliminated, and the time to set up a working service could be reduced". However, the literature does not prescribe a single acceptable implementation approach for RDM, though there are points of convergence that can guide institutions (Henderson & Knott, 2015:50). For instance, Raboin, Reznik-Zellen and Salo (2012: 145) reporting on the lessons, challenges and opportunities from RDM services implementation approaches adopted by three institutions in the USA (Tufts University, Medford, University of Massachusetts, Amherst, University of Wisconsin, Madison) concluded that, securing management commitment and buy-in, capacity of staff to provide RDM services, and the changing needs of researchers were critical issues of concern in any RDM implementation programme. The VCUL experience focused on fostering strong strategic partnerships and active engagement with stakeholders; rigorous educational campaign; taking inventory of existing data and data management resources and facilities; and starting small with easier support services like promoting the DMPTool for writing data management plans through "consultations, group presentations, and teaching opportunities" (Henderson & Knott, 2015:50-53).

A similar study conducted at the University of Sussex revealed that the implementation approach involved engaging with faculties and understanding their RDM needs and practices using three schools (School of Life Sciences, School of Education and Social Work, and School of Media, Film and Music) as representative samples in order to make recommendations to the University management and to provide a focus for developing sustainable data support services at the University (Ball, 2013). These will also provide the foundation to define the roles of the Library, the Computing Unit and the Research Office (Ball, 2013: 260) corroborating the assertion by Jones, Pryor and Whyte (2013: 2-3) when they write that satisfying the research data management needs of a research project is often successful where responsibility is shared across multiple stakeholders including "support and administrative services units [(Library, IT Support, and Research Administration)], and researchers".

Jones (2013:1) reporting on the case of Monash University which avowed that the story of RDM development and uptake at Monash can be linked to the high-level support from its RDM-champion who is the "Deputy Vice-Chancellor for Research [and] Monash's Provost" as well as the strong engagement with Monash's research community. This is consistent with the findings of Henderson and Knott (2015:50-53) and Raboin, Reznik-Zellen and Salo (2012: 145). The approach adopted was a "user-orientated, collaborative approach" which saw the "Library, Monash e-Research Centre (MeRC) and eSolutions (central [IT unit])" collaborating with researchers to "develop and implement an RDM strategy" at Monash University (Jones, 2013:1). Furthermore, the RDM programme at Monash addressed several infrastructure and service components that are consistent with the components of RDM services proposed by Jones, Pryor and Whyte (2013). These are discussed in section 2.3.3. Monash's mantra of "Adopt, Adapt, Develop" distinguishes its approach from other strategies. Regarding this mantra, the University adopts any RDM solution that has already been developed and in use (or being explored), and where necessary adapts this solution to fit the Monash institutional and researcher context (Jones, 2013:4). Developing new solution such as "RDM platforms" will only arise where alternative solutions are unavailable (Jones, 2013:4).

Wilson et al., (2011:275-276) report that the recognition of the complex academic structure and the diverse data management needs across the University of Oxford does not make it feasible to develop the individual components of a coordinated data management infrastructure in isolation and hope that the system works effectively. It is essential therefore to develop an infrastructure to support data management at every stage of the data lifecycles to ensure that the full "potential value" of data is achieved (Wilson et al., 2011:275-276).

Procter, Halfpenny and Voss (2012:142) however disagree, stating that "[a] 'big bang' approach where a fully fledged service is rolled out across an HEI on a given date is not feasible". The reasons cited are: the fact that gathering "requirements across all research

groups and disciplines" in itself is a complex and time-consuming endeavour; the evolving nature of RDM services (Ogier et al., 2014:101) does not make a one-time requirement assessment sufficient to capture subsequent developments; and a onetime roll-out will simply impose a huge financial outlay on HEIs which may not be sustainable given the limited budget allocations available for HEIs (Procter, Halfpenny & Voss, 2012:142). Indeed, the lessons learnt at Monash University corroborate the complex and time consuming nature of RDM implementation (Jones, 2013:6) and the phased approach adopted for the MaDAM project for instance (Poschen et al., 2012; Collins et al., 2010; Goff et al., 2010; Poschen et al., 2010) supports the argument that a progressive rollout would be a much more feasible alternative (Procter, Halfpenny & Voss, 2012:143).

Further case studies conducted by Rans and Jones (2013) at the Universities of Edinburgh, Southampton and Surrey confirm many of these findings and emphasise the point made about the converging and diverging perspectives of RDM implementation. Ultimately, these implementation strategies are underscored by the need to focus data support services on the researcher; his/her needs, participation, practices and disciplinary etiquettes; seek management support and commitment from the inception; and harnessing available resources, facilities and service units to deliver data support for researchers and research teams at every stage of the research lifecycle.

2.3.2. Developing Institutional RDM services and Infrastructure

According to Qin (2013:215-216) the notion of a data (RDM) infrastructure is analogous to "'Infrastructure as a Service (IaaS)', which is a standardized, highly automated offering, where compute resources, complemented by storage and networking capabilities are owned and hosted by a service provider and offered to customers on-demand". The complexities of the data-intensive and networked nature of the new research paradigm [e-Research/e-Science] (Cox and Pinfield, 2014:299; Wang, 2013:1) requires that these infrastructures support extensive collaboration, integration of datasets and even more accessibility across distributed user communities (Whyte, 2012a:209-210). Jones, Pryor and Whyte (2012:143) recognised that infrastructure development is a complex task to undertake because it must address the varied and changing requirements as well as both short and long term implications.

Whyte (2012a:209-210) discussed three levels of infrastructure development; institutional, national and continental (using the European context). Qin (2013:216) agrees stating that infrastructure may be designed at the institutional, research community, national or international levels. More importantly, she opined that there are three dimensions of data infrastructure; technology tools and platforms, data and metadata standards, and institutional data policies which are incorporated into the research enterprise and through which services are developed and implemented (Qin, 2013:215,216). These services are discussed in details in section 2.3.3.

A how-to guide on developing RDM services by Jones, Pryor and Whyte (2013) summarises the implementation strategies and lessons of more than forty UK HEIs and provides practical steps to guide RDM implementation and data service and infrastructure development elsewhere. Particularly for technical infrastructure development, the report on RDM technical infrastructure exploration for the University of Sheffield is an excellent resource (Lewis, 2014). Essentially the development is a three-step process:

i. Acknowledging need for and initiating change;
ii. Conducting audit and assessing requirements; and
iii. [Re]designing and implementing services and infrastructure (Jones, Pryor & Whyte, 2013:4-5; Jones, Pryor & Whyte, 2012:144-145).

2.3.2.1. Acknowledging need for and initiating change

Culture remains a critical factor in the RDM challenge from the institutional level (e.g. governance) to the disciplinary level (e.g. diversity of research). The call by van Deventer and Pienaar (2015:43) for institutions to consider their local and cultural context when implementing RDM, and the assertion by Cox and Pinfield (2014:300) that "technical capabilities, ethical considerations, legal issues and governance frameworks" fundamentally shape how RDM is practised or supported within different institutional contexts gives emphasis to this point. Jones, Pryor and Whyte (2013:4) stated that before data infrastructures and services are selected for implementation, institutions must perform a number of functions such as scope institutional requirements, identify specific data challenges and needs, and justify and build consensus on specific data services and expectations. These functions together with other strategic decisions that the institution must take will be influenced to a large extent by a blend of existing research culture, governance style, "scale of its competitive aspirations" and also by "the measures at its disposal to predict and manage resources within the broader profile of its operating plan" (Jones, Pryor & Whyte, 2013: 4).

It is absolutely necessary to begin by assembling a team of interested stakeholders, led by an RDM champion, preferably a senior administrative official [such as the Pro-Vice-Chancellor (Pro-VC) for Research or "an esteemed researcher with cross-discipline authority"] (Jones, Pryor & Whyte, 2013:4; Jones, Pryor & Whyte, 2012:144). This is essential for securing management commitment and buy-in as noted by Raboin, Reznik-Zellen and Salo (2012: 145). The team must have representation from a broad spectrum of the University community such as the "library, IT services, the research office, ethical advisors, archive services, legal experts and the research community" (Jones, Pryor & Whyte, 2013: 4).

2.3.2.2. Conducting audit and assessing requirements

The next step in the process for developing data infrastructure and services is to conduct a requirement gathering exercise to understand prevailing RDM practices and expectations,

and also do a gap analysis to identify any existing gaps between current capabilities and support, and future or expected capabilities and support (Jones, Pryor & Whyte, 2012:144). These assessments are also essential to develop action plans to meet future expectations (Jones, Pryor & Whyte, 2013: 4).

Several assessment tools exist to support this phase of the data service development but the two prominent ones are the DCC's Data Asset Framework (DAF) for data auditing and Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) for benchmarking. Both tools and their functions complement each other and have been used in various fora to discover critical RDM issues and key areas for action (Jones, Pryor & Whyte, 2013: 4). A much detailed discussion on requirements assessment and available tools is provided under section 2.4.

2.3.2.3. [Re]designing and implementing services

After scoping institutional requirements and identifying specific challenges, weaknesses and needs, institutions can go ahead and develop tailored services to address those challenges and needs. A number of service models and transferable toolkits have been developed such as those from the DCC (http://www.dcc.ac.uk/resources/external/tools-services) and DataONE (https://www.dataone.org/investigator-toolkit) which provide a useful head-start (Jones, Pryor & Whyte, 2013: 4). Whyte and Tedds (2011:3-4) also discussed several RDM programme outputs (tools, templates, platforms and training materials) that have been useful in developing RDM capabilities, skills, services and infrastructure.

In terms of service models, Figure 2.2 provides a graphical visualisation of components, services and processes that were common to the experience of universities developing RDM services through the JISCMRD programmes and "DCC institutional engagements" (Jones, Pryor & Whyte, 2013: 2). These components also correspond roughly with the "five key areas of action" agreed on at a two-day workshop by key stakeholders to forge a joint path for RDM in UK universities over the next five years; these are: "Policy development and implementation"; "Skills and capability"; "Infrastructure and interoperability"; "Incentives for researchers and support stakeholders" and "Business case and sustainability" (Brown, Bruce & Kernohan, 2015:5).

Jones, Pryor and Whyte (2013:5) add that in developing potential services and infrastructure solutions, it is important to engage the body of users and also implement the solution using a pilot approach to ensure a seamless fit within the existing research framework, practices and systems. An across-board deployment can be carried out once sufficient use case scenarios have been observed and tested (Jones, Pryor & Whyte, 2013: 5).

2.3.3. Components of RDM Services and Infrastructure

Supporting effective RDM and data sharing as mandated by the EPSRC policy framework, for instance, requires institutions to develop a coherent strategy and a range of services to

meet the needs of their research communities (Jones, Pryor & Whyte, 2013: 5). Figure 2.2 shows a snapshot of the aspects of RDM infrastructure that require service solutions. In the proceeding paragraphs, each of these components will be discussed, making references to relevant cases where possible and outlining practical guidelines from the literature.

2.3.3.1. RDM policy and strategy

The development of an overarching strategy and policy for RDM are two critical activities that are required (often as first steps) for a coherent data service development. The RDM strategy sets out an institution's RDM expectations and objectives for the future and maps out an action plan to arrive at that future (Pryor, 2014:14; Rans & Jones, 2013:1). To do this, an institution must assess its current position; identify and define future/expected position; and outline a course of action (with key milestones and timelines) to achieve that (Jones, Pryor & Whyte, 2013: 6). The requirement gathering activity discussed in sections 2.3.2.2 and 2.4 is an important component in strategy development.

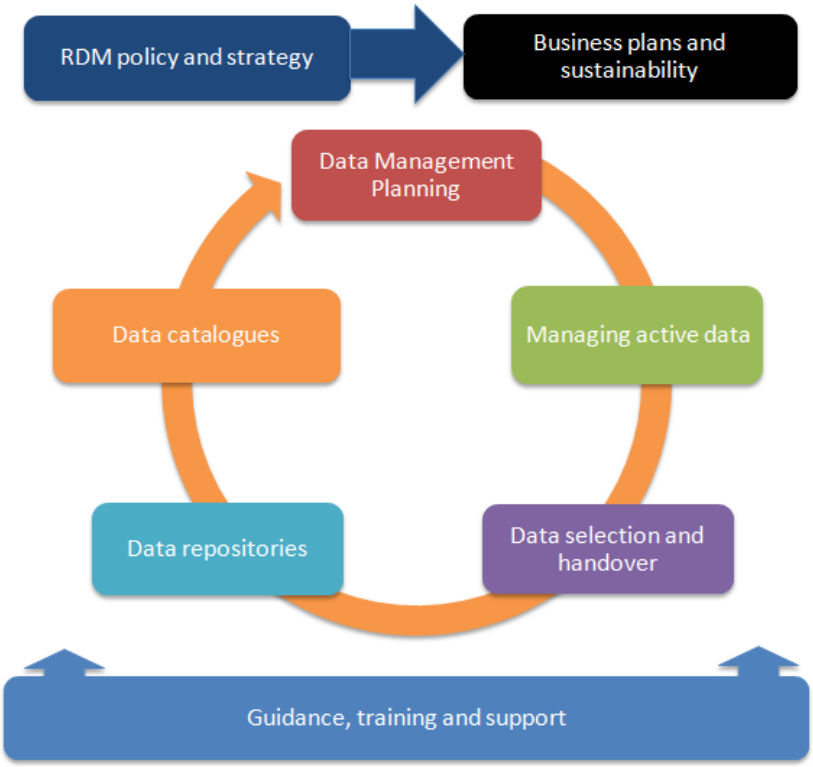


Figure 2.2 Components of an RDM service (reproduced from Jones, Pryor & Whyte, 2013)

The RDM policy on the other hand "describes the principles that an institution has agreed will guide the decisions and the actions necessary to achieve desired outcomes" (Pryor, 2014:16-17; *see also* Rans & Jones, 2013:1). It requires a broad base consultation of all key players within the university community in order to develop a policy that is practical and acceptable (Jones, Pryor & Whyte, 2013: 6). Donnelly (2014) outlines five steps for developing an RDM policy which are; "map out the existing regulatory framework";

"develop a table of contents"; "get managerial backing"; "consult, draft, revise"; and "approve and implement". Jones, Pryor and Whyte (2013: 6) suggest that draft policies should be pitched to a small group of stakeholders such as senior management, researchers or support service staff to ensure that they are comprehensible and address the main issues before being finally ratified. This is consistent with Donnelly's (2014:2) fourth step which involves drafting, consulting on and revising the policy based on tangible suggestions. Policies are best kept brief and in high-level abstractions with the flexibility of being updated regularly to accommodate developments in data services (Jones, Pryor & Whyte, 2013: 6; Rans & Jones, 2013:1). A pilot implementation and incentives may better facilitate uptake (Jones, Pryor & Whyte, 2013: 6). Rans and Jones (2013) provide an overview of three cases on RDM strategy development at the University of Edinburgh, the University of Southampton and the University of Surrey and discuss the emerging lessons and common challenges.

In the UK, the DCC provides a list of institutional RDM roadmaps [strategies based on the EPSRC expectations] (DCC, 2013a) as well as data policies of UK universities (DCC, 2013b) and many overseas examples including that of Monash University in Australia and John Hopkins University in the USA (DCC, 2013c; Beitz, Dharmawardena & Searle, 2012). Horton and DCC (2014) also provide a comparative tabular analysis of UK HEIs' data policies. Jones, Pryor and Whyte (2013: 6-7) further noted that most HEIs implementing RDM begin with data management policy development but the case of the University of Bath which developed a strategy first before dealing with policy as a component of that strategy gives support to the assertion of Pryor (2014:14) that strategy development is perhaps the logical place to start. What is important is to consider the institutional context and culture and adopt an approach that is fit for purpose.

2.3.3.2. Business plan and sustainability

The RDM business plan is basically the reference document that operationalizes the strategy and roadmap for implementing RDM in an institution. It must define the RDM objectives, outline potential costs and expenditures (such as on human resources and infrastructure development), identify "critical success factors", provide an explanation of expected benefits as well as "measures to comply with legal, statutory and funder obligations, and a sustainability and exit plan" (Pryor, 2014:15; Jones, Pryor & Whyte, 2013: 7; Rans & Jones, 2013:1).

Jones, Pryor and Whyte (2013) also outlined some practical guidelines. For instance, due to the uncertain nature of public financing, especially for long-term programmes in HEIs, it will be prudent to stagger implementation "over three, five and ten years [short-medium-long term]"; this three to five year phase-in corresponds roughly to the time period used by many HEIs in planning and executing operating plans (Jones, Pryor & Whyte, 2013: 7). Because of this, it is imperative for RDM business plans to echo the vision and mission of each HEI as stated in its strategic plan. This strategic alignment is also essential for making a

strong business case as well as gaining legitimacy for the business plan (Jones, Pryor & Whyte, 2013: 7).

Again, issues bordering on costing models, funding models, the justification for RDM services and sustainability must also be addressed in the business plan. These have already been discussed in section 2.2.4 but as an addendum some tools and guidelines that may be useful to institutions, particularly those seeking to implement for the first time, include: the "Transparent Approach to Costing (TRAC)" (HEFCE, 2014) and "Full Economic Costing (fEC)" for cost estimation and recovery; and the *KRDS* reports (Beagrie, Lavoie & Woollard, 2010; Beagrie, Chruszcz & Lavoie, 2008) for the cost-benefit justification for RDM infrastructure and data services investments.

Several critical questions must also be answered. For instance, to what extent will managing RDM services be centralised or decentralised [e.g. for "storage or training" services] (Jones, Pryor & Whyte, 2013: 7)? To what extent will it impact on staff roles and competence requirements, especially as RDM becomes a regular university service and an intricate part of the research process (Jones, Pryor & Whyte, 2013: 7)? Answers to these questions and many more will depend partly on the culture of each as well as the scope of their expectations.

2.3.3.3. Guidance, training and support

There are many ways HEIs can provide guidance and training on RDM for their divergent communities. For instance, many HEIs have developed websites that outline best practices and provides links to institutional support as well as useful external resources and tools (Jones, Pryor & Whyte, 2013: 8). This supports the statement by Charbonneau (2013:369) when she writes "[d]ata management training efforts may range from the development of guides, web pages, online tutorials, guidelines, and checklists to presentations and data consultations". A few of the websites have been listed in Table 2.2 in the case of Australian universities. At the University of Pretoria (UP), a subject guide has been developed which provides a link to the University's RDM policy, useful manuals on best practices, tools and templates and a free online training course (MANTRA) (Van Wyk, 2016). The DCC has also compiled a list of RDM guidance websites for the UK and abroad (DCC, n.d.c). Many other useful resources on best practices are also available (Corti et al., 2014; Van den Eynden et al., 2011).

Also, helpdesk systems can be used to route RDM queries (Jones, Pryor & Whyte, 2013: 8). In many other instances, universities provide dedicated email addresses or details of contact persons for RDM support as in the case of UP (http://up-za.beta.libguides.com/c.php?g=356288&p=2428338) and John Hopkins University (http://dmp.data.jhu.edu/). Consultancy services are also useful for personalised support such as getting project-specific answers for RDM planning, using support staff as "sounding

board" to assess the suitability of RDM plans, or providing more technical solutions to researchers (Jones, Pryor & Whyte, 2013: 10-11).

In most cases HEIs also provide RDM training to various groups of target audiences. According to Jones, Pryor and Whyte (2013: 9) these trainings can be categorised under discipline-focused training for researchers, such as those available on the DCC website or as "Continuing Professional Development (CPD)" programmes aimed at re-skilling and empowering support staff (e.g. librarians) to deliver RDM support (DCC, 2016a). Jones, Pryor and Whyte (2013: 9-10) provide a comprehensive tabulation of many of such training and links to their resources.

2.3.3.4. Data Management Planning

The directive from funding agencies for researchers seeking funding for their research to include a data management plan (DMP) as an addendum to their application has been one of the key drivers of RDM in many institutions as discussed in section 2.2.1.2.3. Beyond this, Jones, Pryor and Whyte (2013: 11) add that the growing demand for DMP from HEIs and the many benefits, some of which are discussed in section 2.2.1.1, provide strong justification for data management planning.

To do this, HEIs can adopt one of three approaches;

i. Where HEIs themselves require a DMP, they can develop templates and guides that indicate what information must be included in the DMP. They can also depend on the DCC checklist which synthesises the requirements for DMPs and best practice from a broad spectrum of funder and institutional guidelines under eight categories of questions (DCC, 2013d). Another excellent approach has been to create a compilation of previously successful DMPs which can be consulted by researchers to plan their own data management activity (Jones, Pryor & Whyte, 2013: 11-12).
ii. Make available data management planning tools such as DCC DMP Online (DCC, 2012) or University of California Curation Center (UC3) DMPTool (UC3, 2011). These tools guide researchers to write up their DMPs in accordance with funder requirements, as they do not only ask questions but also provide details of various funder requirements, and facilities for collaborative development and sharing of RDM plans (Jones, Pryor & Whyte, 2013: 12). In some cases, also, these tools may be customised to the specifications and preferences of individual HEIs (Jones, Pryor & Whyte, 2013: 12).
iii. As discussed in section 2.3.2.3, institutions can also provide consultancy and training support on DMP writing.

Responsibility for data management planning according to Jones, Pryor and Whyte (2013:12) is a shared one involving administrative and support services as well as the whole research community within the University. They, however, envisaged that data

management planning skills will become intrinsic to graduate education in the future (Jones, Pryor & Whyte, 2013: 12).

2.3.3.5. Managing active data

Managing storage is a critical aspect of RDM for both researchers and HEIs with two main issues of concern for providing services to support data management during the research active phase; these are:

i. providing adequate space to store the growing quantities of research data from across the target research domain(s); and
ii. availability of suitable applications and systems that make data management and sharing a seamless experience for researchers especially over collaborative and networked environments (Jones, Pryor & Whyte, 2013:13; Jones, Pryor & Whyte, 2012:147).

Several RDM storage infrastructure options and solutions have been suggested and discussed by Jones, Pryor and Whyte (2013). First, it is important to scope institutions' data asset holdings and data management practices (including storage and backups) to understand the extent of the challenge. The requirement gathering activity is again an important step here. According to Jones, Pryor and Whyte (2013: 13), many of such scoping exercises have revealed several handcrafted approaches to storing research data. They added that researchers often preferred to "buy cheap storage and to run their own systems" even when universities provided central storage facilities because these central services were much more expensive. This corroborates the point made by Wilson and Jeffreys (2013:242) stating that it is challenging to convince researchers to use centrally provided infrastructure and services because alternative solutions often give researchers greater control over their data in terms of how processes are managed. However, while these solutions may appear cheaper and much valuable compared to central services, both Jones, Pryor and Whyte (2013: 13) and Wilson and Jeffreys (2013:242) agree that there are often disguised costs and the "risk of data loss and security breach" tends to be significantly higher which could potentially lead to far greater costs in the future.

One solution to this challenge is for HEIs to offer a level of free storage service to their research community such as the model used for data.bris project at the University of Bristol, where a 5 terabyte (TB) capacity of storage is granted by default to every researcher who signs up, at no cost through its dedicated "Research Data Storage Facility (RDSF)". The facility also allows the researcher some level of control in terms of who can access the data and what the lifespan of the data should be. If researchers needed extra storage space above the free 5TB, they were charged on a "Pay Once Store Forever" basis, where "forever" means twenty years (Hiom et al., 2015:479; Jones, Pryor & Whyte, 2013: 13; University of Bristol. n.d.). Jones, Pryor and Whyte (2013: 13) asserted that universities can do this through a variety of approaches such as "[leveraging] their High Performance

Computing (HPC) facilities [as the case for the Sudamih project at the University of Oxford (Wilson et al., 2011)], extending the capacity of existing filestores or exploring secure cloud storage options".

2.3.3.5.1. Cloud storage services

The use of Cloud services for research data storage to a large extent implies lesser investment burden comparative to developing an in-house system from scratch but may pose critical data security challenges over distributed systems (Jones, Pryor & Whyte, 2013: 13) and also pose limitations for collaboration (Whyte, 2012a:224-225). Jones, Pryor and Whyte (2013: 13) suggest that such decision should be made based on a risk-benefit analysis of each case scenario.

2.3.3.5.2. Academic "dropbox" service

Due to some of the issues raised above including "the perceived security and legal risks from using third-party services", institutions are beginning to provide "Dropbox-like" services to their communities. These services will enable researchers to access, backup and sync their data across a range of devices (Jones, Pryor & Whyte, 2013: 14; Rice & Haywood, 2011:238). Others also take the form of "Database as a Service (DaaS)" such as the one developed through the Sudamih project allowing remote access, backup, syncing and collaborative editing of data (Wilson et al., 2011: 285).

2.3.3.5.3. RDM Platforms

In some cases universities have had to develop new RDM systems from scratch such as the case of the University of Manchester where the Manchester Data Management (MaDAM) programme and its Research Infrastructure (RI) team have developed a number of platforms and services to support RDM (http://ri.itservices.manchester.ac.uk/) as well as the Monash University (Jones, Pryor & Whyte, 2013: 14). Jones, Pryor and Whyte (2013: 14) also mentioned two other discipline-specific platforms that support active RDM: OMERO (http://www.openmicroscopy.org/site/products/omero) which is a "client-server software for visualization, management and analysis of biological microscope images" and BRISSKit (https://www.brisskit.le.ac.uk/) which is an "open source Biomedical Research Software as a Service". They intimated that the rhetoric of RDM platforms is shifting toward a "globally-accessible cross-platform" and the optimisation of mobile devices for RDM platform services could be explored (Jones, Pryor & Whyte, 2013: 14).

This study seeks to add National Data service infrastructure as a potential solution to managing active data in HEIs and to supplement what has been suggested by Jones, Pryor and Whyte (2013). This is briefly discussed below with notable examples.

2.3.3.5.4. National Data service infrastructure

According to Collins (2012:168-169) there is a growing collaboration between data centres and HEIs to support RDM as data centres "represent an important and valued asset within

the research infrastructure" and conspicuously play a critical role in supporting researchers and institutional RDM endeavour. Wilson et al., (2011: 278-279) add that the use of such national infrastructure offers better "economies of scale and concentration of expertise".

A number of nationwide data services are emerging which can be explored as new models for supporting and supplementing HEIs' RDM efforts. Examples are:

i. UK Research Data Service [UKRDS] which was a joint project by the Research Libraries UK (RLUK) and the Russel Group of IT Directors (RUGIT), with funding support from the HEFCE, JISC and University Modernisation Fund (UMF), to explore the feasibility of developing and managing a national level research data support and shared service (Whyte, 2012a: 223-224);
ii. UK Data Service which is a service jointly funded by the Economic and Social Research Council (ESRC), University of Essex, University of Manchester and JISC to support researchers to share and re-use high-quality "social and economic data" (UK Data Service, n.d.);
iii. Australia's Research Data Service [RDS], which is a federally funded project through the National Collaborative Research Infrastructure Strategy (NCRIS) programme to provide a national integration of useful research data across nine research domains over shared high capacity storage infrastructure for collaborative access (Australian Government, n.d.); and
iv. Data Intensive Research initiative of South Africa (DIRISA) which is an initiative of the "National Integrated Cyberinfrastructure System (NICIS) to coordinate and promote sound research data management practices and data intensive research" in South Africa (DST, 2013: 25; DIRISA, n.d.).

Other national approaches can be also seen in Germany and the Netherland (Whyte, 2012a: 220).

2.3.3.6. Data Selection and handover

Exercising a data selection decision is an inevitable aspect of RDM, especially because it is practically not feasible to preserve every single item of data that is generated. Among several reasons cited, Whyte and Wilson (2010:2) emphasised the current and particularly recurring future cost of retaining and curating digital data over a long period. Establishing a selection process will therefore be appropriate for prioritising what data to preserve, particularly as HEIs must demonstrate judicious use of public money (Jones, Pryor & Whyte, 2013: 15). To ensure transparency, consistency and accountability in data selection decisions, Whyte and Wilson (2010) outlined seven criteria that must be captured in a data selection policy and for the appraisal of research data.

Jones, Pryor and Whyte (2013: 15) assert that, particularly among research groups where research data sharing is seen as divulging personal and critical intellectual assets, policy directives for handing over data can be threatening; even more so, where decisions must be

made as to which data to retain and which to dispose of. In order for HEIs to encourage their researchers to select and deposit their data, and more importantly, address the challenges associated with establishing procedures and processes for selecting and depositing research data, a four-pronged approach may be adopted:

i. A high-level central support for researchers particularly in understanding the import of decisions about what data will qualify for selection and long-term preservation, as well as what levels of services are available for data that are considered as having re-use value;
ii. Departmental and research group level advocacy and guidance for prioritising data based on the potential to repurpose that data within the research domain in which it was created or in other disciplines;
iii. Making available easy-to-use deposit tools (especially in the area of metadata documentation and deposit workflow) that will encourage researchers to hand over their data; and
iv. Creating a deposit agreement to outline the "terms and conditions" as well as the responsibilities of researchers (depositors) and HEIs (service providers) towards the deposited data. This could alleviate any fear of losing control over their data on the part of researchers and also engender trust between the two parties (Jones, Pryor & Whyte, 2013: 15,16)

2.3.3.7. Data repositories

A data repository is where curated data is kept; data may be designated as being of archival significance and, therefore, to be kept indefinitely or it may be consigned to storage for a predetermined period, following which it is erased. It is, therefore, an integral aspect of the RDM service infrastructure. There are three ways by which HEIs can provide this service; developing an in-house institutional data repository; collaborating with external data repository service providers; and pointing users to available and relevant data services (Jones, Pryor & Whyte, 2013: 15, 16).

2.3.3.7.1. Internally developed data repository

Institutional repositories (IRs) are service suites that an HEI offers to its community including the management, organisation and dissemination of the digital asset collections mainly produced by its faculty members (Halder & Chandra, 2013:1). The IR has been used mostly to store and publish scholarly articles, "but their technical infrastructure can be extended to enable the curation of data without the development or purchase of an entirely new software platform" (Jones, Pryor & Whyte, 2013: 18). Developing a secure data repository infrastructure from the scratch may require a significant capital investment, but for those with an already functional IR hoping to convert or adapt it to support data archiving, there are many useful case studies that can serve as a guide (Jones, Pryor & Whyte, 2013: 18). Wong (2009) reports on such a case from the Hong Kong University of Science and

Technology (HKUST), where the Library was exploring the modalities for hosting datasets on their existing IR.

Again, a number of critical issues must be put into considerations such as skills requirements -- whether there are staff with rudimentary skills who can be retrained to take up curation and RDM support roles, or whether existing skills will suffice -- as well as the range of services to be provided by the repository service (Jones, Pryor & Whyte, 2013: 18).

2.3.3.7.2. External data repository

There are several existing data centres that support the archiving, management and sharing of research data. These established services provide a concentration of expertise and infrastructure that can support HEIs' RDM efforts (Collins, 2012:168-169; Wilson et al., 2011: 278-279) though these may not be bespoke services. Many of these are subject-based or "community-based" data repositories, either developed commercially, or by publishers specifically to archive and make openly accessible data from articles they have published, or as an international initiative (Jones, Pryor & Whyte, 2013: 18). The Registry of Research Data Repositories (http://www.re3data.org/) provides links to a collection of data repositories worldwide, organised under subject, location and content types. There are also funder supported facilities such as UK Data Archive and Natural Environment Research Council (NERC) data centres (http://www.nerc.ac.uk/research/sites/data/) which have strict criteria for accepting data (Jones, Pryor & Whyte, 2013: 19). More importantly, any decision to incorporate data centres into an HEI's RDM strategy must take into consideration among other things their criteria for acceptance, scope of services, charges, and terms and conditions (Jones, Pryor & Whyte, 2013: 19).

2.3.3.7.3. Pointing researchers to relevant services

Where researchers in an institution are not aware of existing data repositories, support services may provide advisory and guidance services in that respect by directing them to related data repositories both within or related to their discipline that are trustworthy and may accept their data (Jones, Pryor & Whyte, 2013: 19). The Registry of Research Data Repositories may be useful resources in providing this service. It is, however, important especially in developing the institutional strategy that, as much as possible, the strategy specifies the extent to which external services must conform to the internal standards implicit in the institutional strategy and criteria on "data preservation and sharing" (Jones, Pryor & Whyte, 2013: 19).

2.3.3.8. Data catalogues

A data catalogue is a record of datasets and the metadata descriptions of the datasets which in some cases may provide direct linkage to accessing the data (Cox, Pinfield & Smith, 2016:9). Both the "RCUK Common Principles on Data Policy" and "EPSRC Expectations" recognise the need for data to be made discoverable through a comprehensive metadata description and "a robust Digital Object Identifier (DOI)", and place a demand on HEIs to

keep a complete record of their research data and make the metadata freely accessible over the internet (Jones, Pryor & Whyte, 2013: 19).

However, most HEIs according to Jones, Pryor and Whyte (2013: 20) are more concerned with;

i. *collect[ing] the metadata in a seamless way, [and] integrating systems wherever possible to avoid placing additional administrative burdens on researchers;* and
ii. *ensur[ing] that standards are followed wherever possible to enable export into any national system as it develops.*

This is to say that institutions exploring data deposit and metadata capturing workflows must look out for "opportunities to automatically harvest data from related systems to avoid re-entry" and even though there is no single acceptable standard for developing catalogues, it is advisable for HEIs to comply with existing standards and initiatives rather than develop new ones (Jones, Pryor & Whyte, 2013: 20). The DataCite metadata schema (http://schema.datacite.org/) provides a useful guide for HEIs and many more related resources are also available on the British Library page (British Library, 2016). It is however interesting to note the nuanced opinion of Whyte and Allard (2014:5) who, in discussing data services development and the disparities in RDM context, state that "[c]atering for a research environment limits the scope to introduce standardised 'enterprise' solutions, as these may not meet the needs of research to explore novel ways of working with data".

2.4. Requirement gathering and need assessment for RDM implementation

According to Henderson and Knott (2015:49) gathering requirements and conducting "need assessment surveys" are critical activities of any RDM implementation initiative. Indeed, this is a truism and has been corroborated by the several RDM programmes and initiatives reported in the literature (Ball, 2013; Rans & Jones 2013; Wilson & Jeffreys, 2013; Raboin, Reznik-Zellen & Salo, 2012; Rice & Haywood, 2011; Wilson et al., 2011; Takeda et al., 2010) that had some component of requirement gathering and assessment. As alluded to early on, such institutional scoping activities help RDM working groups or implementation teams to better appreciate prevailing RDM practices, needs and expectations, capabilities, infrastructure and support as well as strength and weakness which are essential both for strategy and services development (Jones, Pryor & Whyte, 2013: 4; Rans & Jones 2013:1).

Whyte and Allard (2014) discussed the subject extensively, documenting five key approaches and methods for requirement discovery in the RDM domain. These are:

i. Case studies
ii. Surveys
iii. Data Curation Profiles (DCP)
iv. Stakeholder profiles, personas and usage scenarios

v. Development workshops (Whyte & Allard, 2014: 11-19).

They opined that whatever requirement discovery methodology HEIs adopt, it is essential to ensure that it is "regularly reviewed to make sure it can accommodate new challenges of scale and complexity" (Whyte & Allard, 2014:20). For the purpose of this dissertation, however, the survey method will be discussed focusing on the two prominent assessment tools (DAF and CARDIO) and how they have been used to scope institutional requirements. In many cases, RDM initiatives have begun with a requirement gathering exercise using the DAF methodology (Davidson et al., 2014:217), a merger of DAF and CARDIO tools (Pryor, 2013:187) or using both to complement each other (Whyte & Allard, 2014:14; Jones, Pryor & Whyte, 2013: 4).

2.4.1. DAF

The DAF tool was developed in a JISC-funded "Data Audit Framework Development" (DAFD) project at the Humanities Advanced Technology and information Institute (HATII), University of Glasgow, with support from the DCC. (Jones et al., 2009:5; DAF, n.d.). It is a requirement gathering tool for scoping existing data assets as well as RDM practices. DAF is a survey instrument used to "identify, locate, describe and assess how [HEIs' researchers] are managing their research data assets". To put it lucidly, DAF is an assessment tool used to:

a. *find out what data assets are being created and held within institutions;*
b. *explore how those data are stored, managed, shared and reused;*
c. *identify any risks e.g. misuse, data loss or irretrievability;*
d. *learn about researchers' attitudes towards data creation and sharing;*
e. *suggest ways to improve ongoing data management* (DAF, 2009:3).

Auditing institutional data assets and practices using the DAF instrument is a four-phased incremental process (Whyte & Allard, 2014:12; DAF, 2009:5; Jones et al., 2009:11; DAF, n.d.). These are:

Phase 1 – *Planning*:

This involves defining the objectives and scope of the data survey. It may also require the appointment of an auditor (internal or external), securing department's consent and scheduling time and location for the survey. A reconnaissance study may also be conducted to gain acquaintance with the departmental context and possible data assets (Whyte & Allard, 2014:12; DAF, 2009:6; Jones et al., 2009:15; DAF, n.d.).

Phase 2 – *Identifying and classifying existing data assets*:

This involves creating a comprehensive inventory of data assets within a department and categorising them according to their value to the department or focus of the survey. This categorisation will inform the scope of the assessment and subsequent activities. A combination of questionnaire, interview and desk-study is considered the best approach to

elicit information from the research community on existing data assets, their quantities and their location (Whyte & Allard, 2014:12; DAF, 2009:7-8; Jones et al., 2009:22-23; DAF, n.d.).

Phase 3 – *Assessing the management of data assets*:

At this stage, a more focused assessment and detailed analysis is conducted using again the instruments mentioned in phase 2. The goal is to understand existing data practices and whether current institutional support (such as resources and infrastructure) for data "management, control and curation is sufficient to maintain the value of data asset[s]". It may require using the lifecycle model to assess and analyse RDM practices (at each stage) to help identify gaps such as in "data policy and current data creation and curation procedures" (Whyte & Allard, 2014:12; DAF, 2009:8; Jones et al., 2009:31; DAF, n.d.).

Phase 4 – *Reporting findings and making recommendations*:

This final stage involves collating all the responses and observations gathered into a final audit report. The findings in the audit report will inform the kinds of recommendations that are required for improving RDM practices in the institution or department (Whyte & Allard, 2014:12; DAF, 2009:9; Jones et al., 2009:37; DAF, n.d.).

2.4.1.1. A review of some DAF assessments

There is an extensive literature on the use of DAF for data and data practices assessments across several HEIs. Ekmekcioglu and Rice (2009) report on the "Edinburgh Data Audit Framework Implementation" project which tested the DAF tool at five research units (Centre for Integrative Physiology [CIP]; School of Divinity; Economic and Social History; SFC Brain Imaging Research Centre; and Institute for Astronomy) at the University of Edinburgh. The findings of the case studies suggested the need for improvement in RDM practices; for instance, it was observed that data storage was done in a somewhat ad hoc manner with ineffective documentation making it quite difficult to retrieve data, in most cases, even though many of the existing data sets were found to be valuable and hard to regenerate (Ekmekcioglu & Rice, 2009:14-15). They also outlined some challenges including the scope of describing derived data as well as timing especially because the DAF tool was intensive and time-consuming to conduct (Ekmekcioglu & Rice, 2009:15). These corroborate the issues raised by Gibbs (2009:6) in a similar survey at the University of Southampton, admonishing HEIs to always make room for contingency time.

Other DAF projects as reported by Jones, Ball and Ekmekcioglu (2008) shared similar challenges as reported by Ekmekcioglu and Rice (2009) and Gibbs (2009). These included inadequate storage facilities, "lack of formal policies for creating and managing data", ineffective data management practices leading to possible "data corruption and integrity issues", and in many cases lack of technical infrastructure for archiving and long-term preservation (Rice & Haywood, 2011:235; Jones, Ball & Ekmekcioglu, 2008:116-117).

The DAF methodology was also employed to scope data assets and data management practices in the Sudamih Project at the University of Oxford (Meriel, 2010:4-5). While admitting that the small sample size used for the assessment made it difficult to make concrete generalisations, Meriel (2010:5, 9) insist that it provided some useful snapshots. For instance, it shows that there are many existing humanities data assets which may be of value to other researchers but that are not shared for a number of reasons (Meriel, 2010:7-9). The eventual outcomes of the Sudamih project -- Database as a Service (DaaS) infrastructure and RDM training -- provide the opportunity and impetus for sharing research data (Meriel, 2010:11).

Many of these early assessments provided useful lessons and feedback for the DAFD project and for updating the DAF methodology and questionnaire, particularly the time factor, accessing information and issues on scope and level of granularity (Jones et al., 2009:63; Jones, Ball & Ekmekcioglu, 2008:119).

2.4.2. CARDIO

The CARDIO tool (DCC, 2011b) is a benchmarking method for assessing gaps between current and expected institutional RDM support and capabilities (Whyte & Allard, 2014:13), to develop data management strategies, and which is "typically applied at the departmental or research group level" (DCC, 2011a). It has also been used to assess HEIs' "readiness to support [RDM] and curation" (Davidson et al., 2014:217). The tool which was developed as part of a JISC-funded "Integrated Data Management Planning (IDMP) Toolkit and Support" project, "blends key aspects of the Cornell three-legged stool model, DAF, AIDA [Assessing Institutional Digital Assets], DRAMBORA [Digital Repository Audit Method Based On Risk Assessment], and DMP" (DCC, 2011a). CARDIO essentially enables RDM working groups to:

i. *collaboratively assess data management requirements, activity, and capacity at [an] institution;*
ii. *build consensus between data creators, information managers and service providers;*
iii. *identify practical goals for improvement in data management provision and support;*
iv. *identify operational inefficiencies and opportunities for cost saving;*
v. *make a compelling case to senior managers for investment in data management support* (DCC, 2011a).

It is a collaborative assessment tool which is used to engage stakeholders to rate institutional support and capabilities (i.e. infrastructure, staffing, financing and so forth) over different levels of maturity (Jones, 2014: 5; Whyte & Allard, 2014:13). This is done through a questionnaire administered online or face-to-face [typically in a workshop setting] (Whyte & Allard, 2014:14). The goal is to build consensus between "data creators

[researchers], information managers and service providers" by identifying gaps in institutional support and capabilities and negotiating a common focus for improvement in a pragmatic and cost saving manner (Pryor, 2013:187; Knight, 2012:233).

As postulated in the three-legged stool model, CARDIO assesses three elements which are essential for a workable and sustainable RDM effort. These are organisation, technology and resources (Jones, 2014: 8; Whyte & Allard, 2014:13; Pryor, 2013:187; DCC, 2011b). Jones (2014) provides details of the coverage of each of these elements, and they are also illustrated in the CARDIO workflow document available through the DCC website (DCC, 2011a). The CARDIO tool allows for local adaptation (Pryor, 2013:187) and the three "elements can be assessed to different degrees of granularity according to the level of engagement required" (Whyte & Allard, 2014:13). There are three basic ways the CARDIO tool can be used:

a. CARDIO-lite: a quick survey consisting of nine questions with three multiple choice responses for each which is used to gain a snapshot of HEIs' current position (strengths and weaknesses) in terms of RDM support and capabilities (Jones, 2014: 19; Whyte & Allard, 2014:13). The CARDIO short quiz is available online (DCC, 2011b) and also at http://www.dcc.ac.uk/sites/default/files/documents/Dundee-roadshow/Exercise-1-CARDIO-quiz.pdf.
b. Roadmap matrix: developed to benchmark progress towards developing an RDM roadmap which is EPSRC requirement for compliance (Jones, 2014: 22; DCC, 2011a). This version consists of nine questions grouped into three categories ("RDM policy, strategy development, governance and sustainability"; "Data management support and staff development"; and "Research data storage, preservation and sharing") with five categories of responses for each ("envisioning and initiating", "discovering", "designing and piloting", "rolling out", and "embedding") representing the stage of RDM service development (Whyte & Allard, 2014:14; DCC, 2013e). It is designed typically for a workshop scenario and is available through the DCC website (DCC, 2011a).
c. Maturity Assessment: Whyte and Allard (2014:14) state that RDM services may conform to funder expectation but not necessarily be optimal. In that case, the "full CARDIO model" (DCC, 2011c) which outlines a set of thirty statements describing best practices for each of the three legs of assessment under the CARDIO methodology may be used to assess the maturity of institutional services against those best practices (Whyte & Allard, 2014:14).

CARDIO assessment is principally a five-phased workflow which could be condensed or extended depending on an institution's circumstance (Jones, 2014: 14; Whyte & Allard, 2014:13; DCC, 2011b). These five stages, which may simply be referred to as "5Cs", are outlined below:

Commencement: A coordinator conducts an initial assessment to identify RDM contexts and assign ratings for the different support components identified (DCC, 2011b).

Collaboration: Stakeholders across the institutions are invited to individually rate institutional RDM efforts along the three conceptual legs. The focus here is to elicit stakeholders (participants') perception of institutional support and capacity as well as their own levels of responsibility (DCC, 2011b).

Clarification and Consensus: The responses from the individual assessments must be relayed to the group to identify commonalities and discrepancies in perceptions and to build consensus on what is the real situation. For instance, where library services consider their support adequate to facilitate and support good RDM practices but researchers think otherwise or express a lack of awareness of such services (DCC, 2011b).

Conclusion: Where consensus is achieved, identified strengths, weaknesses and risks must be outlined, together with realistic recommendations for improvement (DCC, 2011b).

Commitment: All stakeholders must now implement recommendations. Identified requirements are provided and responsibilities are assigned. For instance, infrastructure is developed where it is non-existent and is required, or librarians are trained to build their capacity to train and advise researchers on RDM where it found to be lacking and required (DCC, 2011b).

2.4.2.1. A review of some CARDIO assessments

Whyte and Allard (2014) report on how CARDIO was used at the University of Warwick in the UK. It involved DCC staff engaging with faculty and support services staff in a series of workshops to scope institutional requirements and to "examine factors governing institutional interest in RDM support, challenges and gaps in capabilities" (Whyte & Allard, 2014:14). Participants were made to rate institutional capabilities along the EPSRC-compliance version categorisations, giving reasons for the assigned ratings and collaboratively building consensus for each (Whyte & Allard, 2014:14). The comments gathered from the workshop were then used to develop a set of interview questions on a range of issues including data management planning, "support and training", and "expectations about services and priorities" and administered to researchers within two pilot groups (Whyte & Allard, 2014:14). Key findings from the interview were subsequently presented in a follow-up workshop where, in collaboration with support services staff, an action plan was developed to facilitate compliance with the "University's Research Data Management Policy" (Whyte & Allard, 2014:14).

Similarly, the CARDIO methodology was used at the University of the West of England (UWE) in its managing research data pilot project (Fowler, 2012). The "Enterprise maturity model" which is essentially an adaptation of the CARDIO tool, was used in the stakeholder engagement stages of the project's "seven-stage roadmap" (which outlined chronologically

the stages of UWE's RDM journey) to gather stakeholders' perception of current capabilities and what their own aspirations for the future were (Fowler, 2012). The "Target operation model" which reflected the "ideal future [of] UWE RDM status" visualises these aspirations (Fowler, 2012). The findings of these assessments will eventually be included in a draft institutional policy together with a strategy that will provide the impetus for institution-wide uptake and management commitment towards infrastructure development (Fowler, 2012).

Pryor (2013:187), writing on the DCC institutional engagement programme and the approaches adopted to develop RDM capabilities across UK's HEIs, confirmed these examples, stating that the CARDIO analyses were particularly instrumental in putting together compelling business cases to senior management to invest in RDM infrastructure and services.

2.5. RDM and the Library

Several scholars have opined that the management of research data is a new area of engagement for the library and in this surrounding is still evolving in terms of its practices and responsibilities (Ogier et al., 2014:101; Corrall, Kennan & Afzal, 2013:655, 666; Corrall, 2012:105). Some scholars have also challenged the library's role in RDM and whether library staff possess the requisite skills set for such a function (Corrall, 2012:106; Wong, 2009:131). Yet, alternative literature justifying the participation of the library often points to the fact that, as custodians and distributors of scholarly outputs (Henty, 2008:2), the management of the data from such scholarship is considered an ordinary extension of the libraries' mandate (Henderson & Knott, 2015:56; Cox & Pinfield, 2014:300; Lewis, 2010:2).

References have been made to emerging data initiatives in libraries, the creation of RDM portfolios in libraries and the publication of books and articles targeted at libraries and information professionals as evidence of the mainstreaming of RDM in the LIS domain (Corrall, Kennan & Afzal, 2013:645). Others have cited the seeming connection between RDM and open access (OA) which librarians have been advocating vigorously (Cox & Pinfield, 2014:300; Corrall, 2012:108) as justification for the library's engagement in RDM.

Corrall (2012:105) asserted that libraries have capitalised on the developments in RDM, such as the requirement for DMPs by funders to forge strategic partnerships and develop services, which has culminated in the development of an emerging body of literature. Cox, Verbaan and Sen (2012) corroborate this assertion emphasising that librarians are well-placed for a leadership role in RDM because "they have good networks within institutions built through liaison activities" (see also Cox & Pinfield, 2014:300 and Lewis, 2010:24). This perhaps is a direct response to the call by scholars such as Lewis (2010:2) and Wong (2009:131) for libraries to collaborate with other research stakeholders, such as the research office and central computing service, to address the research data challenge.

Corrall, Kennan and Afzal (2013:646), reviewing the literature, outlined some ways that libraries are responding to this research data challenge.

RDM is also increasingly becoming a strategic focus for the library. Cox and Pinfield (2014) studied the role of libraries in RDM and how strategic the development of RDM services is to them. Analysing 116 survey responses from across UK higher education and research institutions, they reported that 72% of the respondents said academic libraries in the UK were actively involved in the development of an institutional RDM policy by the end of 2012, 16% were not involved and another 12% did not know if libraries were involved or not (Cox & Pinfield, 2014:304-305). This, notwithstanding, the results show that RDM services were still not fully developed, with HEIs providing limited data support at their libraries (Cox & Pinfield, 2014:306-307). This finding corroborates an earlier study by Corrall, Kennan and Afzal (2013) that sought to understand the changing role of the library and its implication for research support in academia across four countries (Australia, Ireland, New Zealand and UK). They found that very few libraries in the survey sample offered some kind of RDM services (Corrall, Kennan & Afzal, 2013: 654). Both studies however show that RDM was a strategic priority as many are anticipating the implementation of RDM services in the near future (Cox & Pinfield, 2014:307; Corrall, Kennan & Afzal, 2013:655).

As many libraries are contemplating and planning to implement research data services, Christensen-Dalsgaard et al., (2012: 1) recommended ten data management issues that must be the focus of their roles; these include: forging strategic partnerships; creating a portfolio to be managed by Data librarians; metadata standardisation and services; infrastructure including data repositories and research data support services throughout the data life cycle including Data Management Plans (DMPs), data storage, curation, discovery and citation. This section discusses the role of the library, the skills and competence requirements and the training needs for librarians in supporting RDM.

2.5.1. The role of the Library in RDM

Discussions on the role of the library in RDM have been well documented. A review of the literature (Koltay, 2015:4-6, Cox & Pinfield, 2014:302; Charbonneau, 2013:367; Corrall, Kennan & Afzal, 2013:654; Auckland, 2012:16-31; Corrall, 2012:110-120; Cox, Verbaan & Sen, 2012; Lewis, 2010:11-19; Lyon, 2012: 129-130) revealed nineteen ways that libraries can or are participating in the RDM agenda. Table 2.3 summarises the role of the library in RDM as reported in the literature.

While some authors focused on institutional-level roles, where the library provides services to support data management and curation locally (Koltay, 2015; Charbonneau, 2013; Corrall, Kennan & Afzal, 2013; Auckland, 2012; Lyon, 2012), others have emphasised a broader perspective to include national-level advocacy and curriculum development (Cox & Pinfield, 2014; Corrall, 2012; Cox, Verbaan & Sen, 2012; Lewis, 2010). Lewis's (2010) "research data management pyramid for libraries" (see Figure 2.3), as updated by Corrall

(2012), in particular, highlights service areas where the current expertise of librarians can support RDM and curation (Witt, 2008:195) but Cox and Pinfield (2014:302) and Cox, Verbaan and Sen (2012) go on to map each of these roles to both "existing library roles" and the competence required for the particular role.

As shown in Table 2.3 all of the nine authors reviewed identified providing research data advice (including advice on DMPs and data citation, just to mention a few) as a role of the library. This is followed closely by teaching data literacy (identified by eight authors) and developing local curation capacity (identified by seven authors). These are not surprising revelations as these top three roles correspond strongly to already existing or core functions of the library and perhaps represent areas that the library is best equipped to support. For instance, the advisory roles correspond to the "reference and enquiry" services already provided by libraries, the RDM training role corresponds to the library instruction and "information literacy training" that libraries perform, and the curation role relates to the IR and cataloguing (metadata) functions as well as the OA advocacy role of libraries (Cox & Pinfield, 2014:302; Cox, Verbaan & Sen, 2012).

Leading institutional RDM policy development; supporting data deposit, publishing and sharing; promoting RDM awareness; and developing and managing access to data collections are the next prominent roles as identified by five authors for each. The growing influence of libraries in local RDM policy development has been corroborated in the study by Cox and Pinfield (2014:304-305) in the UK which showed that 72% of the respondents said academic libraries were actively involved in the development of an institutional RDM policy as at the end of 2012. Also, managing collections is a core function of libraries and naturally extendable to datasets, and the long tradition of advocacy culture is relevant to policy development and RDM awareness promotion (Cox & Pinfield, 2014:302; Cox, Verbaan & Sen, 2012).

Other roles include data impact assessment, introducing data management skills into undergraduate research work (dissertations), promoting data re-use, and signposting data management expertise within institutions (identified by four authors each); and data auditing and need assessment (identified by three authors), with the remaining seven roles identified by at least one author (See Table 2.3).

To conclude, these roles are a response to the needs of the changing research landscape (Kim, 2013:503) and a testament to how librarians' roles will continue to expand (Cox, Verbaan & Sen, 2012). In the words of Henty (2008:9) "[t]his is an exciting time to be working in libraries as they take up the challenge of new roles and responsibilities".

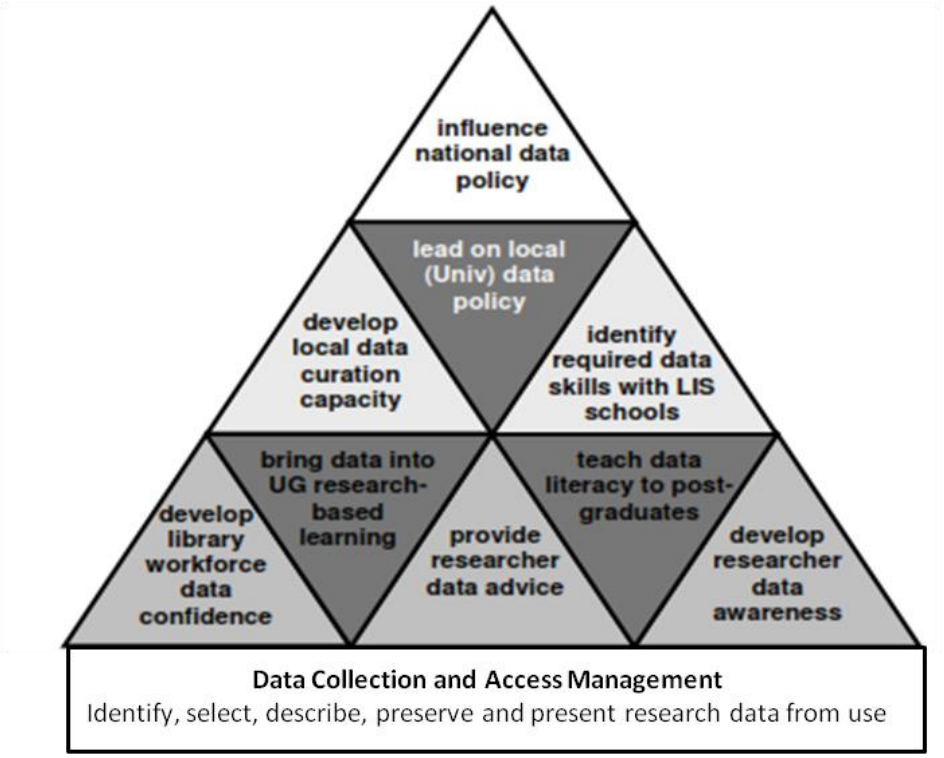


Figure 2.3 "Research data management pyramid for libraries" (reproduced from Corrall, 2012:111; Lewis, 2010:16).

2.5.2. Skill and Competence requirements

The study by Cox and Pinfield revealed that 50% of the respondents did not consider librarians to have the relevant skill-set to manage research data (2014: 308). A third of the respondents also said they believed librarians possessed patches of skills but which may require some extension to be able to meet data services requirements (Cox & Pinfield, 2014:308). These findings are consistent with the assertion by Corrall, Kennan and Afzal (2013:659) when they cited the "need for additional knowledge and skills and […] confidence" for librarians as a constraint on successfully developing research data services. Those who consider librarians to have some relevant skill for RDM often referred to cataloguing and metadata skills among other things (Cox & Pinfield, 2014:308), but Lewis (2010:21) asserted that librarians' knowledge of metadata and cataloguing rules, though relevant, is not enough to make them successful curators of research data.

	Authors / Roles	Koltay (2015)	Cox & Pinfield (2014)	Charbonneau (2013)	Corrall, Kennan & Afzal (2013)	Auckland (2012)	Corrall (2012)	Cox, Verbaan & Sen (2012)	Lewis (2010)	Lyon (2012)	**Total**
1	"Appraising research data"	*								*	2
2	"Assessing data impact (using emerging data metrics)"	*		*		*				*	4
3	"Audit to identify datasets for archiving, create a catalogue of materials or to identify RDM needs"		*	*				*			3
4	"Bring data into undergraduate research-based learning"		*				*	*	*		4
5	"Data mining"					*					1
6	"Develop and managing access to data collections"	*	*				*	*		*	5
7	"Develop LIS workforce data confidence"						*		*		2
8	"Develop local data curation capacity"		*	*	*	*	*	*	*		7
9	"Develop research data awareness (including funder requirements)"		*	*			*	*	*		5
10	"Develop tools to assist researchers to manage their research data"				*						1

11	"Identify required data skills with LIS schools"						*		*		2
12	"Influence national policy"						*		*		2
13	"Lead on local (Institutional Policy)"		*		*		*	*	*		5
14	"Promote data reuse by making known what (relevant dataset) is available internally and externally"		*		*	*		*			4
15	"Provide research data advice (e.g. on data management planning, or on data citations) through formal (web portals) or informal means"	*	*	*	*	*	*	*	*	*	9
16	"Requirement gathering"	*								*	2
17	"Signpost who in the institution should be consulted in relation to a particular question (e.g. on licensing"	*	*					*		*	4
18	"Support data deposit (storage), sharing and publishing"	*		*	*	*				*	5
19	"Teach data literacy (RDM training)"	*	*	*	*		*	*	*	*	8

Table 2.3 Role of Librarians in RDM

Some scholars have investigated the most essential skills and competencies needed by librarians to support RDM (Kahn et al., 2014:302-303; Charbonneau, 2013:367; Corrall, Kennan & Afzal, 2013:663), others such as Cox and Pinfield (2014:302) and Cox, Verbaan and Sen (2012) have gone a step further to map these required competencies to the new roles created by demand for RDM services. Most of the respondents in the survey conducted by Kahn et al. (2014:302) at the LIASA workshop held in collaboration with the DCC identified "database development and software skills" as a critical area to develop for librarians. Corrall, Kennan and Afzal (2013:663) however, identified data curation skills as the most required skill but this was followed closely by "technical and ICT skills". This suggests that ICT skills are very essential for librarians to be able to support RDM effectively.

Other relevant RDM skills and knowledge identified in the literature include knowledge on research and the scientific research process (Charbonneau, 2013:367; Corrall, Kennan & Afzal, 2013:663); principles and best practices in data management and use across disciplines, knowledge of metadata standards (Cox & Pinfield, 2014:302; Charbonneau, 2013:367; Cox, Verbaan & Sen, 2012); knowledge of data life cycles and knowledge of data management tools (Charbonneau, 2013:367).

2.5.3. Education and professional development

Pryor and Donnelly (2009:160) asserted that there are very few data librarians in the UK: these could, most appropriately, be called "accidental" data librarians because they found themselves there by chance and not by design. This is not surprising as their review of academic programmes showed that data management related programmes were almost non-existent, particularly in the UK; examples were found in the USA and Sweden. Lewis (2010:17) stated that the curricula of LIS schools have not been able to keep up with the rate and spate of change in professional practices, justifying the call on "providers of professional education for library and information specialist […] to ensure that the content of their programmes, courses and modules is continually reviewed and updated to anticipate change and reflect developments in the service environment" (Corrall, 2012:120). It is however interesting to find that close to 60% of the total sample in the survey conducted by Corrall, Kennan and Afzal (2013:664-665) preferred RDM to be an elective unit within the LIS curricula rather than a core component.

Incidentally, several HEIs in different countries (including the UK) are beginning to offer courses (usually at postgraduate level) which provide various levels of RDM and curation training [i.e. as full programmes or as an embedded module in another] (Corrall, Kennan & Afzal, 2013:663; Corrall, 2012:120). Examples of such programmes and courses (both academic and non-academic) in countries such as UK, USA, Sweden and Ireland are available on the DCC website (DCC, 2016c) which appears to be an update of Pryor and Donnelly's review. It is also worth adding that in South Africa, the University of Cape Town (UCT) is offering RDM and digital curation as modules in its Masters programme at the Library and Information Studies Centre (LISC) and was expected to offer a taught masters programme

on the subject from 2015 (Kahn et al., 2014; 298). Since September 2015, the UCT Library and Information Studies Centrebegan to run a 3-day short course in RDM for librarians, researchers and data managers. Details are available on the UCT LISC website (http://www.lisc.uct.ac.za/lisc/academic-programme#RDM_Short). This blend of academic programmes and continuous professional development (or on-the-job) training for librarians is essential because as Corrall (2012:120) puts it;

> *[e]ducation and training for data-related library activities needs to be provided for both new professionals and experienced practitioners to enable libraries to fulfil their potential and also to ensure that institutions do not assign responsibilities to others who have relevant subject expertise and/or technology know-how, but lack the informational, managerial and personal abilities that are essential to apply the desired specialist competencies successfully.*

2.6. The EPSRC policy framework on RDM

The EPSRC policy framework (EPSRC, 2011) on RDM spells out nine clear expectations (last updated October 9, 2014) of educational and research institutions towards the management and sharing of EPSRC-funded research data which is underpinned by the RCUK's seven "Common Principles on Data Policy" (RCUK, 2011). These principles are "based on the common theme of access and reuse of research data" (Jones et al., 2015). This policy explicitly mandates research organisations including HEIs to develop appropriate data policies, infrastructure, tools and services to support data management and sharing. These include promoting best practices and ensuring that their [HEIs] researchers, including post-graduate students (mostly involved in a dissertation or thesis work), are aware of these best practices, policies, and also their own individual responsibilities regarding the management and sharing of research data (Wilson & Jeffreys, 2013:241; Jones, 2012a:49). However, the framework is essentially non-prescriptive, giving HEIs the freedom to approach the compliance with the EPSRC expectations in terms of developing policies and practices within their own unique institutional context (Jones et al., 2015; Welgert, 2015).

The framework has also been used in many variants by institutions to scope and benchmark existing RDM services and capabilities or even implement new RDM infrastructure in some cases. Welgert (2015) writes that;

> *[w]hile the policy's focus is on management of EPSRC-funded research data, it has had the wider effect of stimulating universities to think about RDM and the research data lifecycle in a way that goes beyond the requirements of a specific funder*.

The responsibility to ensure effective data management and sharing practices has been most emphasised in HE and research institutions and Jones et al., (2015) affirm that these institutions require established infrastructure and processes to ensure:

i. *[r]etained EPSRC-funded research data preserved for a minimum of ten years*
ii. *[e]ffective data curation is provided throughout full data lifecycle*
iii. *[k]nowledge of publicly-funded research data holdings*
iv. *[d]iscoverability; recording of third party access requests*
v. *[n]otice and justification of access restrictions, for example 'commercially confidential'*
vi. *[a]wareness and use of relevant law, for example FOI [Freedom of information]*
vii. *[a]wareness and compliance with research data policies*
viii. *[a]dequate RDM resource allocation for example from quality-related research(QR) funding or research grants* (see also EPSRC, 2011)

The deadline for full compliance with these expectations was set for May 1, 2015, after which institutions could lose eligibility for EPSRC funding especially where it is evident that they are deliberately and unjustifiable refusing to properly manage and share research data from research funded with public money (Jones et al., 2015). To monitor compliance after the deadline,

> *the EPSRC will ask Pro-VCs for research to complete light-touch self-assessments questionnaires and will start with a dipstick testing after the summer break by checking availability of data underpinning published research for papers published after 1 May 2015* (Jones et al., 2015).

As a result, all institutions receiving EPSRC funding "were expected to develop roadmaps from May 2012" (Jones et al., 2015). The roadmap, which is basically an action plan for meeting the EPSRC expectations, will ideally address three thematic issues which essentially encapsulate the nine-point expectations. These are

i. *Overarching issues on data management policy, strategy, governance and sustainability;*
ii. *Development of support services and increasing RDM capability and skills;*
iii. *Technical infrastructure and services required for storage, preservation and sharing [of research data]* (Jones et al., 2015).

2.6.1. RDM Roadmaps

Rans and Jones (2013:1) define a roadmap as "a basic strategy document that [...] defines areas of proposed activity; […] a high-level work plan" outlining institutional goals and key milestones towards achieving those goals. An RDM roadmap has generally been treated as an essential instrument for delivering compliance with the EPSRC policy and was expected to demonstrate how institutional "policies and processes [would align] with the EPSRC expectations" (Jones et al., 2015).

Pryor (2014:16) asserts that an ideal RDM roadmap should "explain key expectations, current arrangements and gaps in provision; the milestones to be achieved over a specific timeframe to meet expectations and mend those gaps; the roles and responsibilities of individual actors and groups; and the costs". Several of the roadmaps developed by institutions especially in the UK are available via the DCC website (http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps).

Even though the RDM roadmap has been used essentially as a tool to deliver compliance with the EPSRC policy, Pryor (2014:16) posits that it could form the basis for developing an institution's overarching "RDM service strategy" and could therefore represent a key component or possible substitute for the business case. This is consistent with the general perception held by participants at an RDM workshop held at the University of Leeds who considered the roadmap as either an "overarching strategic response" to institutional RDM policy or funder requirements, or as "action-oriented project plans and documents that, taken together with a policy, demonstrate a commitment to change" (Whyte, 2012b). In any case, both perspectives are not contradictory because, as explained by Whyte (2012b), the fact that the EPSRC policy framework encourages HEIs to develop strategies which would align with funders' expectations, and policies that are consistent with the seven common principles of the RCUK, "it should make sense to structure a strategy document according to relevant institutional roles and responsibilities, and the processes or methodology to fulfil them" especially given that the goal is to "align policy and processes with the expectations".

2.7. Summary of chapter and conclusion

In this chapter the growing literature on RDM was reviewed with emphasis on HEIs. The concept of RDM was discussed and the developments on the global and African landscape were examined. In the context of HEIs, the chapter explored the challenges and opportunities for HEIs, implementation strategies, role and responsibilities and competence requirements for the library and infrastructure and support services. A brief discussion of the EPSRC policy framework and RDM roadmaps was also done. These reviews and discussions revealed the following:

i. There seem to be a lack of clarity in the literature on the definition of RDM as a concept and terminology, however, the need for RDM and its relationship with the data and research lifecycles is widely acknowledged.

ii. Effective management and sharing of research data holds value for the researcher, the broader scientific community, academic and research institutions and the public at large.

iii. The RDM agenda is most concentrated in the developed world. Africa's only effort in this area has been concentrated in South Africa which has made tremendous inroads even though it lags behind pacesetters in the developed world.

iv. To develop RDM infrastructure and services, two issues are critical for HEIs: meeting the needs of the varying disciplinary and sub-disciplinary research cultures; and sustainability.

v. There is not a single acceptable implementation approach for RDM, but there are points of convergence and useful lessons to guide institutions.

vi. Requirement gathering and need assessment are critical components (activities) of any RDM implementation initiative.

vii. Securing a high-level support for championing the RDM cause (such as from the Pro-VC for research) is essential for top management commitment and buy-in.

viii. Although a new area of engagement for the library, RDM is increasingly becoming a strategic focus and evolving in terms of its practices and responsibilities. Existing competencies in the libraries are useful but new skills sets need to be acquired to successfully and effectively support RDM.

In the next chapter, the research design and methods applied to the study will be discussed.

CHAPTER THREE: METHODOLOGY

3.1. Introduction

This chapter on methodology addresses the design and the theoretical underpinnings for this research work. It describes the sampling methods, data collection methods and instruments, how data has been analysed, quality control mechanisms and the ethical issues binding this study.

3.2. Research Approach and Design

A research design of a study is the overarching strategy or plan of action that a researcher maps out to perform an empirical inquiry (Pickard, 2008:52, 83). Creswell (2009:3) defines it as the "plans and procedures" that will guide the whole research activity – from general assumptions to specific methods and tools for data collection, analysis, and interpretation. It is essentially a systematic framework that describes how a scientific investigation will be carried out. The choice of research approach or research design adopted for a study is premised on several factors such as the research problem or the research questions that need to be answered, the philosophy or paradigm preference that the researcher brings into the research process, the strategies of inquiry, individual experience, available resources and the audience (Creswell, 2009:3; Pickard, 2008:83).

There are essentially two main research approaches – quantitative and qualitative. According to Creswell (2009:4), quantitative research is used to test "objective theories" by examining the relationships that exist between variables. These variables are measurable and often culminate in statistical analysis and representation of findings. It involves defined assumptions and deductive analysis, building internal checks against bias and ability to replicate findings (Creswell, 2009:4). Qualitative research, on the other hand, is essentially for understanding and exploring how groups or individuals interpret a phenomenon. Data is collected within participants' natural setting, often using multiple methods or sources; data is inductively analysed based on general themes and the derived meanings are subject to the interpretation of the observer (researcher), which makes the researcher an instrument (Leedy & Ormrod, 2014:141-142; Creswell, 2013:44-45; Creswell, 2009:4).

Flick (2011:10-11) states that quantitative research typically begins with a theory from which a hypothesis must be formulated to be empirically tested and the data collection design is standardised for the typically large sample of participants mainly as a strategy to increase the reliability, validity, and objectivity of the study and its outcomes. Qualitative research on the other hand involves selecting few and relevant participants purposively and asking fewer questions which the participants must answer in their own words as a narrative and the data collection design is mostly open and aims at gaining a comprehensive and an in-depth view of the case under study (Flick, 2011:12). It focuses on understanding the subjective meanings of issues from the perspective of participants, taking cognisance of

the latent meanings of a situation as well as the social practices of the participants in their natural habitat in other to discover new elements of a situation under study and to develop theories from these discoveries (Creswell, 2013:45-48; Christensen, Johnson, & Turner, 2011:52; Flick, 2011:12).

What is more, the differences between the two approaches are emphasised at a much deeper level with respect to their paradigmatic stance. Arghode (2012) discussed interpretivism and positivism as the basis for qualitative and quantitative research respectively. Positivists' view of the research world is that of an objective, independent and stable reality that can be discovered and analysed; one in which the researcher (observer) is independent (having no influence on the outcome) of the phenomenon that is being observed; research focus is to determine causality – relationship between variables (i.e. concepts that underlie an observable phenomena and which are measured to determine such relationship); and the aim is to explain how things work (nature of reality) so as to make predictions or control how things happen based on general laws (Pickard, 2008:6-9).

For interpretivists, the research world is socially constructed, where the researcher is embedded in the world being observed and develops subjective meaning through their interactions and experiences with that world (Creswell, 2009:8; Pickard, 2008:12). That is to say that reality (meaning) is varied and multiple rather than universal, and both context and individual perspective are pivotal in shaping that meaning (Pickard, 2008:12). The focus of research is to elicit the views of participants on a situation in their life settings through discourse or observation. Meaning is therefore not arrived at by any statistical procedures but rather the contextual factors that shaped the interaction between the researcher and the phenomenon, and the goal is to gain a rich and deep understanding of the observable fact within a particular context. Thus, a generalisation of the findings can only be made to the extent that both the sending and receiving contexts possess adequate similarities that warrant such transfers (Pickard, 2008:13).

In summary, quantitative research involves quantification and explaining how things work generally; the sample size in such studies is usually large, while qualitative research deals mainly with narratives (words), making sense of the meaning that people assign to a situation within their local context and often necessitates a smaller sample size for better and in-depth exploration of a phenomenon (Christensen, Johnson & Turner, 2011:29-30; Creswell, 2009:3).

The qualitative approach has been adopted as the most suitable approach for this research work because the overarching aim of this investigation is to understand the level of RDM development and preparedness at the University of Ghana (UG) in order to develop a roadmap for implementation. This aim of the study and the research questions do not involve any quantification or measurements (Creswell, 2013:48), rather the researcher is relying on participants' opinion and perception of the institution's level of RDM

development and preparedness which can best be elicited through dialogue or open-ended questions.

3.3. Research strategy

Among the several strategies of enquiry available within qualitative research, a case study strategy was chosen for this research because it was considered as the most appropriate in gaining a "comprehensive and in-depth understanding" of the research problem within the specific context and case and more so because the interpretivist worldview (which is the paradigmatic basis for this study) is most concerned with individual contexts (Pickard, 2008:13, 86).

The lack of a common definition for a case study research has been emphasised in the literature on research methodology (Creswell, 2013: 97; Pickard, 2008:85; Gerring, 2007:17). However, the definitions proffered by Christensen, Johnson and Turner, (2011:374) and Gerring (2007:20) suggest that a case study research focuses on comprehensively investigating a phenomenon within a defined boundary. Yin's (2009) discussion of the method also brings to fore another perspective, which is, the "empirical investigation of contemporary phenomenon [in a] real-life context". In other words, a case study is more appropriate in examining or exploring contemporary issues or events within specific and bounded context to gain a comprehensive understanding of the case. Creswell (2009:177) had pointed out that the case study strategy is also appropriate in exploring processes, activities or events.

Accordingly, this research focuses on RDM – an emerging area of scientific inquiry and a new area of engagement for academic institutions – which is still evolving in terms of its practices and responsibilities in our part of the world. The study investigates a single case – the University of Ghana – gathering data from a multiplicity of sources including the administrative and support services (library, research office, and institutional ICT directorate) and research community. These have been identified in the literature as major players in defining, developing and implementing RDM at the institutional level (Jones, Pryor & Whyte, 2013:2-3).

3.4. Population and Sampling method

The population encompasses all the people about whom inferences or generalisations will be made based on the study outcomes; it is the larger target group from which the sample is selected particularly because it is impractical or expensive to conduct the study on the population (Christensen, Johnson, & Turner, 2011:150; Pickard, 2008:59, 60). Sampling "is a process of selecting [a smaller] representative" number of people within a larger population to conduct an empirical study (Pickard, 2008:59). The outcome is a sample that "is a subset of the population" (Christensen, Johnson, & Turner, 2011:150: Sproull, 2002:109). Two broad sampling methods can be found in the literature – probability and non-probability sampling or random and non-random sampling. Probability sampling methods provide

"every member of the population an equal chance" of selection as a sample because the selection is done randomly. It is most appropriate for quantitative research because of the larger samples sizes involved and the need for generalization of findings; it also has a lesser risk of bias compared to non-random methods (Leedy & Ormrod, 2014:220-221, 154; Christensen, Johnson, & Turner, 2011:150; Pickard, 2008:61-65; Sproull, 2002:112-120).

However, non-random sampling methods are considered most appropriate for qualitative research because qualitative inquiries tend to focus on in-depth investigation of one or few cases "rather than a broad" study of many cases (Creswell, 2013:100; Christensen, Johnson, & Turner, 2011:162). More often than not, purposive sampling is used and even though it has been criticized for its high tendency towards bias and risk of selecting samples that are not representative of the population (Christensen, Johnson, & Turner, 2011:158; Sproull, 2002:118), Patton (2015:264) emphasised that

> *[t]he logic and power of purposeful sampling lies in selecting information-rich cases for in-depth study. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the inquiry…*

This research is based on a qualitative case study approach which focuses on eliciting participants' views to understand the level of RDM development and preparedness at the University of Ghana (UG). The non-probability, purposive sampling method was adopted to educe detailed and useful information from information-rich units and persons within the University about the state of the art and preparedness of RDM development at UG. The target population for this study is a heterogeneous one comprising all faculty (academic) members, students, researchers, and staff of administrative and support service units at UG. However, the study population will be limited to only senior academic members and senior members (management staff) within the three main support service units (Library, IT services, and Research Office).

Because of the heterogeneity of the study population and the data requirements for answering the research questions, a mixture of sampling techniques was used in selecting the samples based on the approaches discussed by Pickard (2008:64). The first group of respondents were selected through "priori criteria sampling" by outlining a set of criteria that must be met. These criteria provided a baseline for identifying purposively information-rich participants within the three service units to answer questions on RDM capabilities, infrastructure and policies. To qualify for selection into the sample, staff must be a senior member and should have worked in that capacity for not less than three years. The researcher believes this provides ample time for the respondent to possess rich information about the capabilities, programmes and policies of the University to be able to respond to the questions.

The second group of respondents were selected through snowball sampling. This was done by working with the Research Office to identify senior researchers within the University, who have not only received external funding for research before but have also done research extensively and published not less than ten scholarly works. Discussions were held with these respondents on their RDM practices (briefly), expectations and perceptions of current RDM capabilities, infrastructure and support at the University of Ghana.

As is often the case for qualitative research, the case study approach typically focuses on studying a few samples of respondents for a detailed understanding of a phenomenon mostly not to generalise the outcomes, but to shed light on a phenomenon within a specific context (Creswell, 2013:157). Keeping in line with the logic of purposive sampling, and to avoid the risk of duplication, a total of seven participants were selected for this study. An initial sample of five had been planned but this had to be increased to the current number of seven because there were aspects of the interaction that were referred to other participants that had rich information on those aspects. Table 3.1 illustrates the breakdown of participants selected for this study. This sample size is roughly consistent with what is recommended by Creswell (2013:157) who asserted that a sample size in the region of five is appropriate for a single case study research.

Respondents' Code	Category	Criteria
R1	UGCS	A minimum of three years working in the capacity of senior member at the UGCS
R2	Library	A minimum of three years working in the capacity of senior member in the library
R3	Library	A minimum of three years working in the capacity of senior member in the library
R4	UGCS	A minimum of three years working in the capacity of senior member in the library
R5	ORID	A minimum of three years working in the capacity of middle-level management
R6	Researcher – Department of Botany	Senior researchers of UG, a receiver of external funding through the ORID, extensive research, at least ten published scholarly works
R7	Researcher – School of Pharmacy	Senior researchers of UG, a receiver of external funding through the ORID, extensive research, at least ten published scholarly works
Total	**7**	

Table 3.1 Breakdown of participants selected into study sample

3.5. Data Collection methods and instruments

Both primary and secondary data were collected and analysed for this study. Primary data represents original data that is collected by the researcher from the field, by observing or

interacting with the study population (Gelo, Braakmann, & Benetka, 2008:275). This data was useful for gaining in-depth understanding of institutional capabilities, requirements and preparedness and was collected using semi-structured interviews.

Secondary data on the other hand are those data gathered from books, journal articles, official documents to mention a few and constitutes data that was generated by another person, that has been reported or published in such media (Bhattacherjee, 2012:39; Gelo, Braakmann, & Benetka, 2008:275). This data source was used to complement and supplement data from primary sources and includes institutional documents such as policies, strategies and administrative documents that are accessible. This also provided useful background information that was used in formulating the interview questions as well as identifying priority areas and developing the roadmap for RDM.

What is more, the rationale for using both data sources was to establish patterns through triangulation. Triangulation is when researchers employ multiple methods or sources of data to corroborate evidence, highlight convergence in the data or provide a better and deeper appreciation of the phenomenon (Flick, 2014: 30; Leedy & Ormrod, 2014:104,106; Creswell, 2013:251; Christensen, Johnson, & Turner, 2011:53). Patton (2015:317) however noted that rather than just being used to corroborate findings, the rationale for triangulating is to test for consistency in data and since different enquiry approaches are susceptible to different real-life nuances, understanding emerging variations in findings across the various data sources used can be illuminating. This makes triangulation an effective strategy for increasing the validity of the study findings (Creswell, 2013:251).

Finally, the decision to combine data collection methods and sources is consistent with the practices for qualitative case study research as espoused in the extant literature on qualitative inquiry (Creswell, 2013; Christensen, Johnson, & Turner, 2011; Pickard, 2008).

3.5.1. Interviews

According to Pickard (2008:171) interviews are the most used data collection method in library and information science (LIS) research and is the most appropriate technique for qualitative and in-depth studies such as case study research. It is defined as a data collection method where by the interviewer (typically the researcher) questions people (interviewees) to discover inter alia their opinions, experiences, perspectives, knowledge, values and expectations, usually in a face to face interaction, though it could also be a technology-mediated interaction such as a phone interview (Christensen, Johnson, & Turner, 2011:56; Sproull, 2002:162). Some of the strengths highlighted in the literature include the ability to probe and ask follow-up questions, opportunity to ask for clarifications and its appropriateness for eliciting in-depth information about a phenomenon including latent contextual details. Yet the method has been criticized for being costly, time-consuming and may produce inaccurate data due to reactive effects or selective recall from the interviewees (Christensen, Johnson, & Turner, 2011:58; Pickard, 2008:171-173; Sproull,

2002:162-165). The researcher believes that the triangulation of data collection methods and sources will help minimize these weaknesses.

With regards to the type of interview to be conducted, the semi-structured interview was chosen. There are other types – structured and unstructured interviews (Pickard, 2008:175) – however, Leedy and Ormrod (2014:156) avowed that interviews in qualitative studies are seldom structured. The choice of a semi-structured approach was mainly because it affords the researcher flexibility compared to a structured approach and at the same time focuses on the main issues which can help save time and avoid redundant information compared to an unstructured approach (Flick, 2011:112-113). An interview guide is a necessary instrument for this approach and must make room for probing and discourse on the thematic areas of the interview (Flick, 2011:112). According to Patton (2015:439), "[the] interview guide lists the questions or issues that are to be explored in the course of an interview" and as an advantage, allows the researcher to maximize use of time. An interview guide was therefore developed for all categories of respondents to reflect broadly the research questions (see Appendix A).

3.5.2. Document analysis

As a secondary data source, document analysis involves the examination of existing documents such as institutional records including memoranda, correspondence, official publications, and reports or datasets produced by another person or for other purposes that have relevance for the phenomenon under study. Its strengths lie in its obtrusiveness and unlikelihood of a reactive effect and the fact that it can serve as useful source for discovering contextual and background information as well as corroborating and verifying data from other collection methods, but it might provide incomplete information on the phenomenon (Patton, 2015:14; Christensen, Johnson, & Turner, 2011:60-61; Flick, 2011:122; Sproull, 2002:170-172).

In this study, document analysis was used as a complementary and supplementary method to gain insight into the state of the art of RDM development and preparedness at UG. Documents examined include institutional policies and guidelines (mostly research and data management related), strategies and administrative documents that are accessible.

3.6. Data analysis

Leedy and Ormrod (2014:143-144) reviewing literature opine that the steps involved in a typical qualitative case study data analysis are: "organisation of details about the case", "categorization of data", "interpretation of single instances", "identification of patterns", and "synthesis and generalizations". It is also important that in a single case study (where multiple participants will be interviewed), each case (participant response) is intensively analysed as a separate entity before concatenating them to find patterns or disparities (Patton, 2015:536; Christensen, Johnson & Turner, 2011:376).

Considering the nature of data collected for this study, a qualitative data analysis was used, specifically, thematic analysis (Flick, 2014:421; Flick, 2011:152). This was considered more appropriate for this study design as it allowed the researcher to develop a thematic structure from the data collected. The thematic structure was used to compare responses from the different participants to see where there is consensus or disagreement on an area of institutional capability, infrastructure and support (Flick, 2011:153).

The analysis process begun with a playback of the interview recordings and making notes on major topics and key terms that were mentioned in the interviews starting with the first interview; only specific portions that were considered useful for direct quotation were transcribed into a text file. These topics were then grouped into categories through a coding process using colours and annotations.

The next step was to crosscheck the categories within the individual interviews conducted to identify commonalities and patterns in the categories or topics; these were regrouped and condensed to form themes which were then discussed. Throughout the analysis process, the researcher continued to look out for the categories developed from the first case in the subsequent cases, but modified and updated them as and when new or contradictory aspects emerged, so as to analyse the views of those participants who form part of a particular interpretation (Flick, 2011:153).

3.7. Ethical consideration

Adherence to ethical conduct is an important component of social research, especially when research involves human subjects. Ethics in research aims at ensuring that researchers conduct their studies in a way that is transparent and not injurious to participants in any way (Flick, 2014: 48-50; Christensen, Johnson & Turner, 2011:96; Flick, 2011:216). It is absolutely necessary for researchers to take cognizance of the ethical ramifications of their study and conduct during the study. Several authors have discussed the main ethical issues that must be observed (Flick, 2014; Leedy & Ormrod, 2014; Christensen, Johnson & Turner, 2011; Flick, 2011; Pickard, 2008).

Accordingly, and as a general rule to scientific inquiry at the University of Pretoria, the following ethical considerations have guided this study:

1. The participants willingly participated in the study.
2. Participants were informed about the purpose of the study and assured about their anonymity before they gave their consent for conducting the interviews.
3. Information about participants was handled with strict confidentiality and their anonymity was ensured during the analysis and representation of data

4. Ethical clearance was sought for all data collection instruments and for carrying out this study from the Faculty of Engineering Built Environment and Information Technology (EBIT) Research Ethics Committee.

5. Permission was sought from authorities of the University of Ghana to conduct this study (See Appendix B).

3.8. Trustworthiness of Research

According to Creswell (2009: 190), validity and reliability in qualitative research do not carry the same connotation as in quantitative research. Nevertheless, using such terms as criteria for establishing the value of qualitative research has been criticized by interpretivist researchers as importing positivists' language which is incompatible with qualitative inquiry (Creswell, 2013:246; Pickard, 2008:18). Accordingly, a different set of criteria – credibility, transferability, dependability, and confirmability – have been proposed as the alternative equivalent for validity and reliability (Creswell, 2013:246-253; Pickard, 2008:18-21) and are considered more appropriate for this qualitative case study design.

3.8.1. Credibility

Credibility is the interpretivist's equivalence of internal validation and focuses on demonstrating persuasiveness in the evidence that is generated from the inquiry (Creswell, 2013:246). Common strategies for this criterion include prolonged engagement with research participants, triangulation, and respondent validation (member checking) (Creswell, 2013:250-252; Pickard, 2008:20). In this study, triangulation of data collection techniques and sources was employed.

3.8.2. Transferability

This criterion also mirrors external validation or the extent of generalisation in quantitative research. The goal, however, in qualitative research is the transferability of empirical evidence which is dependent on contextual proximity rather than a blanket generalisation (Pickard, 2008:20). Hence, a thick description of the context (University of Ghana) has been provided as part of chapter one to provide pertinent background information about the case study and to provide a standard for transferability (Creswell, 2013:252; Pickard, 2008:20).

3.8.3. Dependability

Creswell (2013:246) stated that rather than reliability, qualitative researchers strive for dependability which focuses on the "[accuracy] of the research process" and the appropriateness of the methods used. Typically, external auditors are recruited for the appraisal (Creswell, 2013:252; Pickard, 2008:20). While it has been pointed out that the method and instruments used for gathering data for this study have a probability of bias, the researcher still considers it as most appropriate in the context of this study which seeks

to understand the level of RDM development and preparedness at UG. What is more, internal strategies like triangulation of data methods and sources are in place to mitigate such risks. This criterion, however, will not be applied in its entirety to prevent the risk of exposing the privacy and confidentiality of study participants.

3.8.4. Confirmability

Interpretivist researchers strive for confirmability rather than objectivity in establishing the value of data (Creswell, 2013:246). The goal is to ensure limited researcher bias as well as establish that the study findings (which are the subjective interpretations of the researcher) were based on the data collected from the field and not just the researcher's theoretical proclivities and research interest (Pickard, 2008:21). Though the data (audio recordings and partially transcribed interviews) for this study have not been attached to the final work, it will be made available to anyone who needs it, but in a way that does not expose the identity of the participants. A data management plan (DMP) has been included as an appendix to this chapter which clarifies issues of data access and interpretation (see Appendix C).

With regards to the subjectivity of this study, Creswell (2013:246) stated that clarifying researcher bias from the start is a strategy that helps readers to put the study in context and to understand the researcher's perspective, bias or any assumption that may have impacted the interpretation and approach to the inquiry. In any case, scholars like Patton (2015), Leedy and Ormrod (2014), Creswell (2013) and Pickard (2008) have all made the point that research in itself is inherently subjective, whether with regards to paradigmatic stance or design or the instruments used for gathering the data, and throughout the work these have been acknowledged and clarified.

3.9. Conclusion

Chapter three discussed the methodology, the design and the theoretical underpinnings for the study. It also described the sampling methods, data collection methods and instruments, how data was analysed, the ethical issues binding this study and the trustworthiness of this study. The next chapter, chapter four, will present the findings gathered through the data collection activities and a discussion of those findings.

CHAPTER FOUR: ANALYSIS OF FINDINGS AND DISCUSSION

4.1. Introduction

The goal of this study is to assess the state of the art of RDM development and institutional preparedness at UG through existing data management activities and capabilities. In this chapter, the findings garnered from interviewing respondents and analysing relevant institutional policies and documents are analysed, presented and discussed. In accordance with the criteria described in Chapter Three *Methodology*, the researcher identified patterns of responses and themes that have relevance to the research objectives of the study. These have been categorized and presented as research findings under the following sub–headings; 1) RDM awareness and institutional response; 2) Institutional capabilities; 3) Current activities and practices, and 4) Requirements and other issues.

4.2. Analysis of Findings

4.2.1. RDM awareness and institutional response

The section presents findings on general conceptions of RDM, awareness of the risks associated with poor data management, institutional attitudes, and policy issues.

Generally, respondents demonstrated a vague understanding of what constitute RDM. Most of the respondents, however, demonstrated some appreciable level of understanding of the need for RDM but not exactly what it is. For instance, one respondent conjectured that RDM could be synonymous with "research management", but all respondents outlined a number of critical issues that show that they appreciate the need for RDM. Indeed, this appreciation is further highlighted in the introductory section of the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management":

> *"The primary role of data management is to ensure the highest possible degree of integrity, reliability, and continuity in research. It also affords some level of institutional memory".*

The closest response to what constitute RDM as most respondents indicated was data storage and so many of the supports highlighted were focused on storage capabilities and services.

> *Another way to look at it is a way for us to keep for them in one place what they could easily lose, so that whenever they want it they can get access to it* [R3].

> *So I think really that's where the University should go, having a platform like that where datasets for research can be stored* [R4].

> *…so we need to build some sort of a repository or data management system and back it up with a policy so that every research that emanates from the system, the data is deposit in the repository like we do for the repository for publication of research articles*…[R5].

4.2.1.1. Awareness of risks associated with poor data management

There is a general appreciation of the risks associated with poor data management. The most common risk that came up among the responses was the risk of losing of data. This is to be expected, particularly, as the researchers had indicated that they needed the data to write their paper. Other risks that came up from the interviews were waste of time, energy, and money.

> *When you lose your data it means the time, money, energy spent is gone, and the purpose of research is to try to solve a problem, so you have not solved or resolve an issue if the data is lost. If anything, you have caused financial loss* [R6].

> *For me, these days, people are moving from hard copy notebooks to electronic and I see a big risk, because you wake up one day and the computer is not waking up, and if you don't have a backup then it's gone* [R7].

These issues are also reemphasised in the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" which states:

> *"If the data are not properly protected, the investment, whether public or private, could become worthless. The responsible handling of data begins with proper storage and protection from accidental damage, loss, or theft".*

Furthermore, one respondent noted the potential institutional risk that poor research data management and lack of institutional infrastructure poses to the University, admonishing that the University takes a proactive stance on RDM.

> *So now everything seems ok, but one day suppose somebody comes and say this nice data you generated let's see your data repository and it's not there and we say it's [Mr 'A'] who's keeping it on his computer and he is nowhere to be found, or what is the evidence that he is not even doctoring them [the data] now and then? I think there may be issues, so maybe the earlier we look at it the better. And people have doctored information, not from here, abroad, so it can happen* [R7].

4.2.1.2. Institutional attitude and position on RDM

Much of the institution's position and attitude towards RDM could be gleaned from its research related policies and specifically, guidelines on record keeping and data

management which will be discussed in details in the succeeding section. Some respondents also gave responses that corroborate and reinforce some of these policy provisions. The data shows that the University is fully interested in an RDM culture especially in the areas of safeguarding institutional data and sharing data among faculty members of the University.

> *…as far as sharing of data and management of data is concerned, I think the University is fully interested* [R5].

This was further highlighted by one respondent who indicated the commitment and interest of both the Vice-Chancellor (who was at the end of his tenure) and the incoming Vice-Chancellor towards the development of an HPC cluster to aid some collaborative research at the Department of Physics.

> *…like I said, it had to be done within [the] shortest possible time, maybe two or three months because the VC then and incoming VC were all interested and we were able to work with procurement to get some few servers put together heavy processor base and memory base to be able to enable them [to] run that thing* [R1].

Adherence to good practices in RDM is a key policy principle in the "UG Research Policy", and the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" spells out many of such best practices.

As far as developing an institutional infrastructure to support RDM is concerned, one respondent noted that the University was far from what is desirable.

> *…I don't think we've reached there yet, we have not built capacity for storage of data as compared to other institution where they have a repository where every research that goes on, the data that is gathered is put in the repository for access by other researchers* [R5].

One researcher also mentioned a lack of an institutional framework to support RDM.

> *[At] Noguchi,…we thought at one point the institute for that matter the University should have control on data but we couldn't get far because there was no institutional framework to back it, [so] people did not really cooperate* [R7].

These opinions notwithstanding, the data showed that the University recognises the need to develop systems such as a "meta-database of research materials/ data repository" to support RDM as stated in the "UG Research Policy". The data also shows that the University is yet to develop such a "meta-database" or any RDM-specific platform and for which reason it is currently unable to fully and comprehensively identify existing institutional data assets, their volumes, and locations.

> *Sincerely, no, I don't think so because… if we don't have the system for capturing this information there is no way the University can know what one researcher is doing, what data they are collecting or has collected, we don't have such systems and so frankly speaking we have no idea the research data that have been collected…* [R5].

But as the data indicates, the lack of institutional infrastructure and systems to support RDM is not expected to be long-term. The University's renewed commitment to research development as enshrined in its new "2014-2024 Strategic Plan" has been activated by the establishment of the Office for Research Innovation and Development (ORID). The ORID under the leadership of a Pro-VC (Research and Development) has over the years developed systems and initiated policies and guidelines to support research at UG. However, owing to resource limitations, and institutional priorities and exigencies, several other systems have been developed while RDM still remains on the agenda.

> *Setting up a research office is a gradual process; you have to look at certain system that you will need immediately and then you build on that, so we have identified ethics, intellectual property and other systems which we've managed to put in place in this short period and it's working, so hopefully maybe the next port of call will be to set up what the policy has said. I think the whole issue is one of prioritization; you can't do everything looking at the limited resources in terms of human resources* [R5].

Finally, the data also indicated that writing a data management plan was currently not an integral part of the requirement for internal funding. Neither is there any indication that the University was planning to embed data management planning into the research process at UG.

4.2.1.3. Institutional policy and guideline on RDM

The data indicates that there is currently no specific RDM policy at UG. The University's policy response to the subject of RDM is captured in section 5.6 of the "UG Research Policy" as one of its key policy statements. The Research policy which was ratified by the Council of the University of Ghana (highest decision-making body at UG) in November, 2012, also addresses other critical areas such as Ethics, Intellectual Property, Academic Freedom and Creativity to mention a few. The policy statement on RDM essentially captures four issues in brief;

a. recognition of RDM as a good research practice and integrity issue:

> *"The credibility of research findings depend[s] on record keeping and good data management";*

"Data management is one of the essential areas of responsible conduct of research".

b. institutional commitment to developing systems to support RDM;

"the University will create a meta-database of research materials/ data repositories"

c. researchers' responsibility as the main steward of research data; and

"Under normal circumstances the original materials and data sets will be held by the PI who undertook the research."

d. mandate of researchers to keep datasets for not less than ten years after the completion of a research project.

"The PI is expected to maintain this data set for a minimum of ten years after the final project close-out. In certain special circumstances, this minimum period may be extended."

The most pronounced institutional document on the subject of RDM is the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management". The guideline which was approved by the ORID Management Board is a much detailed and pragmatic working document which covers many critical aspects of RDM including data ownership, data collection and documentation, data storage and retention, data protection, data privacy and data sharing and publication.

However, one respondent noted that the guideline, unlike the research policy, is not necessarily binding as it only aims to advise and provide guidance to researchers on best practices.

…but a guideline is just to guide you, a policy reinforces or is binding, a guideline is not binding [R5].

Regarding how researchers are made aware of research related policies and guidelines, the data shows that, essentially, policies and guidelines are published on the ORID website and the links are also sent to the university community via an institutional emailing platform.

We send links to faculty members so they know that these policies and guidelines are in existence. What we will be doing going forward is to do a compilation of all our guidelines and policies into a booklet form and then distribute it, making it available all over the place or circulate it to faculty members for them to be aware of it. But of course, we are in an internet age so we expect that people will go to the website instead of the hardcopy [R5].

Interestingly, the two researcher-participants indicated that they were not aware of any institution policies or guidelines on RDM.

> *I have no idea about institutional policy on RDM but I know a research policy exist and then there is ethics policy. I am not particularly aware of any RDM guideline…* [R7].

> *I don't know about a policy, but I have heard you can deposit your paper into a repository. I don't know whether it exist, it might exit but I don't know* [R6].

One respondent also gave an indication of how researchers are monitored to ensure compliance with existing institutional policies and guidelines.

> *…we give grants to researchers…the application is online, there are rules and guidelines for you to follow to access the money, so we can use that as a determinant to see if you are following or not,…so we can use that to determine whether [the researchers] understand what the policies, rules, and guidelines are as far as accessing research funds is concerned. It is the same approach in the case of external funders* [R5].

4.2.2. Current activities and practices

In this section, findings on existing institutional support services and researcher practices that were identified as being germane to RDM at UG are presented.

Responding to questions on current activities and services, respondents expressed varied opinions on the current state of affairs. It is important to note that there are two main categories of respondents in this study, first, the service providers comprising of the Library, ICT Services, and Research Office, and then the data originators (researchers). Responses from the service providers on the subject of current services were partly influenced by how participants' units engage with researchers as well as their own perception of what constitutes support for RDM. For some, support for RDM was inferred, and was considered as part of the general research support they offer; while others said RDM services were simply not currently being offered. For instance, one participant [R1] talked about how the UGCS had been called upon to provide specific services and facilities to aid researchers to complete their research task, using key phrases like "provide some ad-hoc support" and "specialized IT infrastructure [for researchers] to do a specific task" to indicate this. However, another participant discussing the issue from the library's perspective stated that an RDM service was yet to be developed:

> *Unfortunately, that has not come up, we are looking more at the output – the eventual published articles…rather than the raw data that has been collected and I don't think the library has really thought of how that could be managed as a library…* [R3].

Interestingly, one respondent indicated that the idea of RDM has come up for discussion among some librarians within the library.

> *The discussion about datasets has really been within the library, discussions with some librarians who have thought about, can we really start taking it? Should we push the agenda* [R4]?

What is obvious is that the library has not taken any strategic approach to the subject of RDM yet.

While the library's engagement with researchers has mostly been in providing access to information and instruction, it is instructive to also point out as the data shows that the library provides software applications that researchers can use, such as for data analysis. One respondent in the library pointed to this:

> *Apart from the resources that we have, we also have this package for data analysis; the NVivo, for example, is for qualitative data analysis* [R2].

The two researchers who participated in the study also hinted at some limited levels of support. The data also revealed some particulars in researchers' practices and attitude. In all, the data gathered does not point to any formalized institutional services for supporting RDM, but they reveal some activities and practices that represent potentials that can be harnessed.

4.2.2.1. Institutional support

4.2.2.1.1. Support for collaborative research

The responses reveal a culture of collaboration in research at the University of Ghana as well as an institutional commitment to foster and promote such collaborations. One respondent summarized this commitment in relation to the setup and role of the UGCS:

> *…when there is a call for us to provide any kind of support to aid collaboration and research we are always available and we've been doing so…I think that, all in all, the units [IT services, IT planning and Security, and IT infrastructure] have been [working] together to support the faculty, the lecturers, [and] the researchers to collaborate* [R1].

The data also revealed specific support for collaborative research which one respondent eloquently described with vivid examples such as developing HPC facilities and providing storage.

> *There was recently some Italian…collaboration they did with the department of Physics and Computer Science and we had to go in to provide…what we call High Performance Computing infrastructure (HPC). We didn't have that*

infrastructure; we had to build some basic infrastructure to support the kind of lab they wanted to run that project [R1].

…[At the] Institute of African studies….they were doing a collaborative research with [some universities in] USA and I think Canada, there again they needed storage to probably take care of all this and we went in to provide some storage [R1].

4.2.2.1.2. Support for data analysis and computational science

The data reveals some level of research computing support currently in existence at UG through the UGCS. One respondent discussed in details how the UGCS has been very instrumental in building and supporting the HPC facilities for high-speed computational science and data storage at UG, as well as the commitment of the University to develop even more powerful capabilities.

But currently to we are also collaborating with Prof. Awandare and his team at Biochemistry department under the [College of Basic and Applied Sciences], they have had the need to build specialised […] HPC to run genome, I mean to run the kind of multiple algorithms to support the kind of chemical related or biomedical related research they are doing. And that we have collaborated with them and IBM to provide such infrastructure. Even though we have an HPC, it's not so much unique for the rest of the community, so we are trying to build another one that could be leveraged by all other departments… like I said, it had to be done within [the] shortest possible time, maybe two or three months because the VC then and incoming VC were all interested and we were able to work with procurement to get some few servers put together heavy processor base and memory base to be able to enable them [to] run that thing [R1].

As pointed out already, the library also provides data analysis packages for researchers who may need them and so support for data analysis is not only limited to quantitative data but also qualitative data.

4.2.2.1.3. Training and guidance

The responses did not reveal any specific RDM training currently being offered and indeed, this was eloquently expressed by both researchers who participated in the study, responding with phrases like "Never before" [R6] and "Not here, not in the USA" [R7] to show that they have never received any RDM related training. It was, however, instructive to discover guidance on funding application and best practices on data management. A number of institutional documents highlighted these and some have been corroborated by some interviewees. The main institutional document that has explicitly dealt with the

subject of RDM and is thus most germane to the topic is the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management". This document has been put together after intensive and extensive research on best practices and ratified by the ORID Management Board. Among other things, the guideline informs researchers on the need to understand and comply with funder regulations. One respondent describing the research support role of the ORID emphasized that the Research Office provides information to researchers about funding opportunities and requirements:

> *…In terms of providing research grants to researchers to conduct research, ORID plays a vital role…We provide research advisory role, we send information to faculty members regarding funding opportunities through a platform we call 'research alert' just so faculty will know the information as far as research grant is concerned* [R5].

The guideline also provides advice on other aspects of RDM, including data collection and documentation, storage, security, retention and sharing best practices. For instance, the guideline outlines the minimum documentation or metadata that must be provided for every dataset.

> *"Record anything that seems relevant to the project, its data, and the standards of the project. At a minimum, records should include the following information: Date and time; Names and roles of any team members who worked with the data; Materials, instruments, and software used; Identification number(s) to indicate the subject and/or session; Data from the experiment and any pertinent observations from the data's collection. It may also be helpful to include a summary of the day's data collection activities and a task list for the next day."*

> *"Data should be retained for a reasonable period of time to allow other researchers to check results or to use the data for other purposes. There is, however, no common definition of a reasonable period of time."*

> *"Data should however not be shared without the permission of the University."*

With regards to training, the data shows that training researchers to use available tools and applications and improving their own research practices are already embedded in the functions of the service providers at UG. One respondent stated that the UGCS train users on using specific software and on data handling.

> *We have a training unit in UGCS,…that does training for faculty for specific software to use for their research activity, SPSS and the like and then when faculty [members] have issues with how to handle such data […] they always come here and there are people who have been supporting them over the*

years. I think we have a blend of research skills as and when faculty come for us we will make them available [R1].

Furthermore, two respondents mentioned Author Workshops which are essentially organised by Elsevier and facilitated by the Library to expose researchers to all they need to know about authorship including requirements for successful publishing, journal selection, and publisher requirements, as well as impact factors evaluation.

We've also done a bit of training in using research software like reference managers, and then we have also collaborated with publishers and done author workshops for researchers just to enhance the research process. And this has brought out issues of where they could be publishing, how they can select journals for publishing and what is required of an author in publishing [R3].

They are taken through how to… find appropriate journals to publish, ...everything about authorship…The last training we had they [Elsevier] took lecturers through impact factor training [R2].

4.2.2.1.4. Data storage and Publishing

The data shows that some limited storage support is currently in existence at UG, especially through the UGCS which has been providing storage support for some research endeavours at UG. The existing HPC facility, Institutional repository, and data centre infrastructure with huge combined storage capacity is being harnessed by the University to provide storage support through the UGCS and the Library. One respondent illustrated this with examples of projects where the UGCS has been called upon to provide storage space to meet the needs of researchers and avows that part of this storage is used for actively managing and storing research data.

…we've started some kind of data from the HPC programme that I mentioned, they have built data…that are running on these servers and so they are already managing them on those servers. And again our HP cloud matrix is already running and so the data that have been captured or processed or worked on are already there and people are using them [R1].

…we don't have direct control over research data, but we have given them platform to generate the data and save it secured, and manage it...What I mean is we give them storage access to resources that they [need] and we give them access to some software they require, they use it and generate their data and they manage it... [R1].

The story is quite different from the Library where despite having the infrastructure (institutional repository) to support data deposit and publishing, inadequate staffing and

lack of drive seem to be the hindrance. As one respondent illustrated, the current IR policy in principle permits the depositing of research data but this is yet to be actualised.

> *Per the IR policy it is written that it permits the deposit of datasets but in actual fact, we are not at the moment accepting datasets. The system itself can accept datasets, but […] I will tell you point blank that for now based on the staff number that we have with regard to the whole IR setup, I don't think we would be able to immediately start it [supporting data storage/publishing]… so if we are going to do this then we'll need dedicated people, [who will] go for these datasets and make sure they put them in an organised format and submit them on the platform* [R4].

> *There is no challenge, as we speak we can accept it, but for now, everybody […] has tuned their mind that we should make the end products [research paper] available in the IR for people to access, eventually, some day we will accept the raw data also* [R4].

One researcher recounted his experience working with one of the flagship research centres of UG where as far back of 1998 the institute had set up a central storage facility for storing and securing research data files:

> *We tried at Noguchi to put up a data storage system. The institute had a server and the idea was that you can sit in your office and save information on the server and as and when you need it you can retrieve... our interest was first of all the institute having control over data and also safeguarding individuals from losing data* [R7].

Another researcher talked about the existence of a herbarium – an institutional setup where specimens or samples of species collected from the field were deposited for display [R6].

4.2.2.2. Ethical matters

As far as ethics is concerned one respondent emphasised that the current ethical system at UG focuses on the impact that researchers' work will have on the subjects or informants from whom data is gathered and not necessarily the broader issues of RDM.

> *We look at your research not being injurious to either the health of the person or tarnishing the image or reputation of the individual. But as far as where you store your data and others currently it's beyond the mandate of the ethics committee, may be another body or system may be put in place to look at how actual data that are being collected are deposited and managed and used towards the benefit of the research institution* [R5].

However, the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" encourages all research works to be approved by the University [through appropriate ethics committees] so as to identify any special data support need:

> *"Every research work should have the approval of the University so that some protection of data can be offered when necessary".*

4.2.2.3. Researchers' RDM practices

The data reveals some divergent but also common grounds on issues relating to researchers' approach and attitude to research data and institutional RDM support. The researcher-participants in this study were selected from different disciplinary backgrounds – Botany and Pharmacy. The analysis shows that generally the management of research data is done by the researchers themselves. This is also accentuated in the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management":

> *"Most of the specific tasks of data management fall to the PI and Research Director"*

Some common research data types that were picked up from the responses include spreadsheets, laboratory notebooks and observations, test responses, field notebooks, diaries, specimens and samples [R6, R7]. Whilst these data types may seem common to both disciples, the approach to its management varies considerably. Another issue with data type has to do with digital and non-digital or hand written data. In principle, both researchers agree that digitisation has enormous benefit for keeping data safe and making it more accessible but one interviewee cautioned about total reliance on digital tools and platforms.

> *Now with [digitisation] we actually image them so they can be on the web, so if you want to see a specific sample from Ghana you don't have to come here because we can put it on the [internet] for people to see. But that is just the starting now we are trying to make it available online* [R6].

> *For me, these days people are moving from hard copy notebooks to electronic and I see a big risk, because you wake up one day and the computer is not waking up, and if you don't have a backup then it's gone …I do remember my data from a particular study, I forgot, it extinguished itself because now we don't have a floppy disk and that data is on a floppy disk, it's there but I can't use it. So it has discarded itself* [R7].

In all, both researchers agree that data is an important asset for them; however, it appears only to the point where they have published a paper out of it.

> *Without my data I cannot get my article, so data is very important just that once I get the article, the data can go into the background* [R7].

> *Normally, I try to convert every data into a paper the raw data will be there for some time and them I will discard it. But because I have put that into a paper, the paper is there forever. I consider ever research output very important because the data also is what results into the paper* [R6].

4.2.2.3.1. Data publishing and sharing

Regarding their attitude towards data sharing, one researcher showed reluctance referring to data as a "trade secret", but was open to sharing where the researcher had greater control of how and with whom data was shared especially after the researcher had published a paper from the data.

> *So even depositing your DNA things there, it's not that you are very happy about it* [R6].

> *…but the raw data that I used in writing the paper that one I won't give it to anybody….Yes, what I want to share I will share but when I want to share, I have to determine that… I prefer I am able to store the thing myself until if someone needs it and If can think I can give it fine, but just to leave it so that somebody else decides who can have access to it, it's problematic* [R6].

Some respondents speaking from the point of view of service providers noted that unless there is a conscious institutional mandate to share research data, researchers will be reluctant to do so:

> *You know when you go outside people are willing to give out [their] datasets so somebody could look at it and understand exactly what the researcher did or maybe possibly also use it for another research… But [here]…they might hold on to their dataset and not want to share it, unless of course, it is a mandatory thing from the very top, from the vice-chancellor… but if that is not done, I don't think people will be bordered to or interested to keep it on our IR, they will not even think about it* [R4].

> *But again we are looking at the policy direction, if we have a policy that says it is binding on you as a researcher, once you publish in any research outlet put your research [data]…*[R5].

Another researcher said the kind of data generated in the researcher's field will have no use if shared with another person:

> *In my kind of work it is difficult to share data because…, if I want to know whether plant medicine used to treat diabetes works, I have to set my*

> *controls [when conducting the experiment] so that becomes specific, if I give that information to you it doesn't help, you can't say…you are going to use my control and you are going to look at another animal, because the ages of the animal may be different, the species may be different, most of the time the data is exclusive you can't use it* [R7].

4.2.2.3.2. Data storage

One respondent stated that the current practice is that researchers generally store their own research data which puts the University at risk of losing valuable institutional asset.

> *…so you analyse and keep it on your computer, if it is hard copy you save them in files and then you publish….so everybody keeps their data and unfortunately if they die that is the end of the data and the University as far as I am aware cannot get to it* [R7].

This practice was corroborated by another researcher who expressed his preference for storing his own research data.

> *"So I prefer I am able to store the thing myself"* [R6].

The data also revealed that the issue of trust was a critical factor in researchers' attitude towards institutional support for data storage and not only data sharing. It was clear that researchers required guarantees that stored data in a central infrastructure will not be divulged within the period of embargo and that the embargo period will be sufficient to have published from that data.

> *But I suspect maybe part of the problem may be lack of trust, people may think if they probably [store it themselves] nobody can go into my database*…[R7].

> *No, because I don't trust human beings…I know when we do these molecular things [DNA] they will tell you, you must deposit your data in a repository… they tell you that this thing that you are depositing here we will guarantee that no one will see it but after two years it will become public. So if I think that within two years I will be able to write a paper, fine* [R6].

Commenting on whether researchers express interest in depositing data in the University's IR, one respondent from the library stated that no such interests have been shown as the researchers themselves are more interested in depositing their finish research work rather than the raw data into the IR:

> *Well, I haven't really spoken to any researcher who has shown the interest or drive of having his or her dataset within the IR. As I said earlier on, they are also looking at having just the end product of their research there for now […]*

> *so that people will know what they are doing. So they haven't shown any interest to me personally and even to the core team that manages the IR…[R4].*

Institutional support for data storage was however seen as a contingency effort as one respondent intimated:

> *So we thought, not that the institutional database or server won't crash but at least if it's on your computer and on the institute [server], the institute one will serve as a backup for you* [R7].

4.2.2.3.3. Data management planning

The responses also show that none of the researchers have had the need to write a data management plan (DMP) and there is currently no support for writing DPM here at UG. What was also evident in the data is that while writing a DMP is not a requirement for accessing UG funding, one respondent mentioned that some international grant awarding bodies may require DMP. One researcher did recall a question relating to how long data will be kept while applying for funding.

> *Not really, it's more interested in how you are going to execute, analyse the data, the outcomes and how you are going to disseminate the [outcome]. Normally that's what I have seen, but there may be a question "how long you will keep the data?"* [R7].

> *I have not had the need to write an RDM plan* [R6].

> *…I think it is the researchers themselves who do that, not our unit or the University* [R5].

> *Now there is a consciousness because international donors want systems in place, so they require those things, so international donors I believe will require for data management plans but as far as our internal grants are concerned I have not seen that on our forms yet*...[R5].

4.2.2.3.4. Data retention and disposal

Responding to how researchers determine which data is of long-term value and how long data is kept, it was evident that they had no specific retention or discard policy and the decision to keep data was more an issue of personal practice.

> *The kind of data we collect they are small, not population-based data, so I don't have any discard policy… Maybe it's a personal practice, I believe in keeping, how much more my notebook. I have never considered throwing this*

data away, I don't know whether it is right or wrong it a personal thing I don't throw away [R7].

4.2.3. Institutional capabilities

This section touches on technology capabilities, knowledge and skills, and resources (funding) that can be harnessed to support RDM.

There is general mixed feelings about the capability of the University to support RDM even though there is a general consensus that the University should be supporting RDM. Some respondents [R1, R3, R4] from the point of view of service providers feel this could be done sustainably but will require necessary retooling, relevant policies in place and institutional commitment. There is a general lack of personnel with adequate knowledge and skills in RDM that can support researchers. Technology-wise a number of facilities, software and systems were identified that may be harnessed for RDM, this ranges from an HPC facility, a private cloud data centre facility (HP Cloud Matrix) with a vast storage capacity, a functional institutional repository developed with DSpace software and that has the capability to accept datasets and harvest metadata, a robust network infrastructure, internal security systems, external cloud services through collaboration with Google and specific data analysis packages like NVivo and SPSS.

4.2.3.1. Knowledge and skills for RDM

There is a general consensus that support staff do not possess adequate knowledge and skill to support RDM at UG. However, once the relevant staff and units are trained they will be able to acquire the necessary knowledge and support RDM.

I have a feeling that yes most people should have a fair idea, but again this is not a system that has been introduced fully in the University, once that is fully introduced, staff will be trained on how to do things, but I have a feeling that people have a fair idea about how [research] data should be managed [R5].

I don't think so; I think we need to retool. Because in the first place, when we look at the library school from which many of us are trained we don't even talk about those things at all, so we really don't have what it takes,[but] I think we can learn. We don't have the right skills [and tools] now, but we are capable of doing it so long as we are retooled, I think there is so much out there that we can read, and learn, and maybe visit people who are doing it and be exposed to what is being done, we can come and replicate it here [R3].

I don't think currently, [we have the necessary skills to support data curation, but] data curation is not nuclear science so I believe the staff concerned, once we are told to move in that direction within a short period of time people will be abreast with the skills to do that, but then also I believe there should be

> *some formal training in that respect…Curation will have to come in, so I believe even if the person has some informal training I believe at the end of the day there should be some formal training before the person is mandated to do that task* [R4].

One respondent, however, mentioned that to be able to support RDM at UG, staff need to be proficient in ICT and possess good communications skills.

> *We need to be completely computer literate to be able to take up this role… I think a good skill will be good communications skills because we need to approach the researchers and help them to appreciate what it is that we hope to do* [R3].

4.2.3.2. IT infrastructure

There is a general confidence in the systems of the University to support the diverse and growing amount of research data that may exist at the University. Though the data does not reveal any specific RDM platform there was great potential for data storage, preservation and publishing. Another key institutional document that mentions RDM is the "UG Research Policy". The policy spells out the need for UG to develop systems for "acquiring and storing research materials and data in a robust and confidential manner". One respondent attested to the robustness of current IT infrastructure to support RDM:

> *So all I can say is that our infrastructure is capable and robust to keep [data], in fact, it has the capacity to scale so if we've had to increase them, it has the facility to be able to upgrade*...[R1].

4.2.3.2.1. Data storage, preservation and backup

Responding on the capability to support active storage and backing up of data while research is ongoing, it was gathered from one respondent that there was an ongoing collaboration between UG and Google which allows UG to take advantage of the massive storage capability of Google across the globe and to move UG students' data onto that platform. This facility allows for the students' mail systems to be hosted on Google servers while allowing students to access free cloud service like Google drive which provides, even more, storage space than the traditional 15GB free space for Google subscribers. This facility allows students to take advantage of the cloud capabilities such as storing, backing up and syncing their data on the platforms. What is more, staff could also subscribe to the free Google service and access free storage and cloud services.

> *….but we have been smart to move students to the Google cloud which is virtually limitless storage, Google has given us gigabytes and we can always ask for more, in fact, we have that clause that we can negotiate for more* [R1].

> *And that was a fantastic thing, virtually we are not contributing a penny, we are advertising Google, and we are using the platform, our students are getting access to store data which is accessible anywhere on earth where they have access to internet, and as well as also giving us a room to be able to scale our resources for staff, because staff are more stable* [R1].

Apart from the Google cloud service, it was clear that UG had a cloud infrastructure with enormous storage capacity which could be harnessed for RDM purpose. One respondent mentioned an "HP Cloud matrix" facility which has Storage Area Network (SAN) and Network Installation Management (NIM) capabilities:

> *As for storage we have about 120 Terabyte on the cloud infrastructure…we have what we call the HP Cloud matrix...and it comprises of Storage Area Network (SAN), Network Installation Management (NIM), optical devices and multiple storage to really provide the needs of other departments* [R1].

This also provides the University with the capability to support institutional research data backup. But as revealed by one of the respondents, however, such services currently are not provided as wholesale support.

> *We take back-ups of the kind of data that we need to take backup, so when there is an emergency we can restore them. But we do it specifically for platforms that we control*…[R1].

The data shows that UG has a great potential to support data archiving particularly in terms of storage space. It reveals that HP Cloud Matrix which has been described as *"one of the best data centres in all universities in Ghana"* [R1] together with other internal storage capabilities places the University in a pole position to store research data for long-term but then there must be institutional commitment to make this sustainable.

> *Well, currently we are positioned to be able to [support an increasing need for storage space for data], but it's not us sitting in UGCS making that call, the senior management, Vice-Chancellor, Pro-VC, the deans, and heads of departments must all agree that if this is what we are going to do then the right resources are put in there* [R1].

> *As for storage, we have about 120 Terabyte on the cloud infrastructure, and with the other servers that have internal storage, if you add up all those we will be moving into about 150 Terabyte of storage within the University* [R1].

Despite the potential to harness existing infrastructure as well as extend existing file stores to support research data storage and preservation, one missing ingredient as discussed above is the lack of adequate skills to support such a venture. As one respondent puts it directly:

> *I don't think currently, [we have adequate skills to support data curation* [R4].

Respondents also point to the need for a policy to ensure that the decision to support long-term storage of research data can be done sustainably.

> *…if we start giving them storage we must have [a] policy on how long we are keeping the digital data, otherwise, within one year we would have generated so much in terabytes* [R1].

4.2.3.2.2. Network infrastructure and security

One respondent gave a general picture of the current network capability and security arrangements at UG, emphasising that a concerted awareness creation effort was needed to ensure that the systems remained secure.

> *We have reasonable bandwidth, until recently we had 670mbs now we have asked Vodafone to increase it to 1G which is 1000mbs* [R1].

> *We have a first line of security, the network is interfaced with a firewall and … also at the end of the users we have also put other security devices like Universal Threat Management, which is Sophos UTM, we have also put another device that also actually checks within the network to see whether there is any harmful thing coming in – intrusion detection – which has been deployed by the IT Security Unit of the UGCS and that warns us to watch out as the network people provide the necessary remedial solutions…[but] It is a continuous kind of education and appreciation of our environment that will actually secure all of us. That is very relevant. Gradual training and awareness creation for the users will help us secure our data*…[R1].

With respect to systems to support data storage and access, some respondents noted that where storage facilities were provided for any research venture, they have been designed to forbid any unauthorised access to data:

> *…and we have provisioned that the right people who need to have access, have access to this data*...[R1]

> *It was that simple, I can have my folder and no one can access it except the IT manager who maybe can go in…But it was secured, for example, sometimes I'll be [on] my computer and you will see people are trying to access but they are denied, so if you don't grant the permission nobody can go in there* [R7].

4.2.3.2.3. Tools and systems for data analysis

As pointed out early on, there is some degree of departmental level support for computational research. One respondent spoke about an HPC facility for high-speed

computational science and data storage developed for the department of Physics, and current effort to build one for the department of Biochemistry, Cell, and Molecular Biology, as well as the ambition of the University to grow such capabilities to be utilised by the whole university community.

> *But currently to we are also collaborating with Prof. Awandare and his team at Biochemistry department under the [College of Basic and Applied Sciences], they have had the need to build specialized […] HPC to run genome, I mean to run the kind of multiple algorithms to support the kind of chemical related or biomedical related research they are doing…Even though we have an HPC, it's not so much unique for the rest of the community, so we are trying to build another one that could be leveraged by all other departments…* [R1].

Other analysis tools are also available for use. In fact, some respondents spoke about specific software for data analysis and on support for data handling.

> *Apart from the resources that we have, we also have this package for data analysis; the NVivo, for example, is for qualitative data analysis* [R2].

> *We have a training unit in UGCS,…that does training for faculty for specific software to use for their research activity, SPSS and the like and then when faculty [members] have issues with how to handle such data […] they always come here and there are people who have been supporting them over the years* [R1].

4.2.3.2.4. Metadata

One respondent spoke frequently about the capability of the IR to harvest metadata from other systems as well as other manual approaches that can be used to harvest metadata about researchers' publications.

> *We also struck this agreement with BioMed where all publication authored by faculty and researchers from University of Ghana are automatically pushed onto our platform together with the full text. It is something that we are planning to expand, so may be for instance with Elsevier or some other publishers we can have publications authored by our lectures, especially for Elsevier we might probably not get the full text but just the metadata automatically deposited onto our platform and we might then move it to the appropriate collection or something like that*…[R4].

> *…DATAD is able to harvest these things onto their platform so we have that interoperability with other systems where we can harvest things from other people and other can also harvest from us* [R4].

> *For instance, [for] publications of our lecturers, we could generate .csv files from Scopus where we can have access to these metadata and then upload them onto the IR. We did it some few times but we felt it will be best if it is done through the formalised approach where the lecturers say this is my publication put it online, but not us going about harvesting publications of our lecturers from different platforms and putting them on the IR* [R4].

4.2.3.3. Funding RDM

Another issue that frequently came up in the responses was the issue of funding for RDM. Respondents expressed mixed opinions as to whether the University had the wherewithal to develop and implement a successful RDM infrastructure and service. One respondent felt that once the University has been able to fund the development of other systems, it should be able to do so for RDM if the need for an RDM infrastructure was identified.

> *Of cause, so far they [University] provide funding for the systems we have in place, we make annual budgets and the University supports it, so whatever systems we put in place, the University should be able to provide the support* [R5].

This notwithstanding, one respondent hinted at the inadequacy of the University's support for such a venture.

> *…there's always limited amount of budget to support different activities required for the University to run* [R1].

Indeed, one researcher [R6] bemoaned the lack of adequate institutional funding for research and calls for increased efforts and commitment, to improve the pool of research grants available for research.

4.2.4. Requirements and other issues

This section presents data on aspects of RDM identified by respondents as needed and other concomitant issues such as responsibility and sustainability.

Generally, respondents affirmed the importance of RDM and the need for the University to take an active role in its development. This was emphasised with phrases like "It [RDM] is highly necessary" and "it [RDM] is very important". A number of RDM areas and issues came up quite strongly for consideration and these are enumerated below.

4.2.4.1. Policy

Evidently, the most advocated pressure point for consideration, most respondents consistently emphasised the need for an institutional policy on RDM to spell out clearly the University's position, expectations and to drive compliance.

> *We can have a policy and a system in place on how UG data should be captured, managed and shared, I think it will be of a great benefit to the University* [R5].

> *…if we start giving them storage we must have policy on how long we are keeping the digital data, otherwise, within one year we would have generated so much in terabytes…it's a matter of having the right policies so that you know that when students generated data when they leave, you clear it and make it available for another fresh batch of students* [R1].

> *And maybe have a policy whether we are storing for 5 years, again it should be [back up] because anybody can destroy information and if we say we are storing 5 years then nobody has any business destroying information within that period* [R7].

> *…so we need to build some sort of a repository or data management system and back it up with a policy so that every research that emanates from the system, the data is deposit in the repository* [R5].

4.2.4.2. Storage

One respondent spoke about the need for institutional storage facilities whether offered as a centralised facilitates or departmental-based.

> *But as at now, I believe it's a weakness in the university set up because there is no central point where the employees are expected to put their data…* [R7].

> *…so that even if not at a central place, each school or department should have a data system to collect data and protect it, whether it's small, specific [or] exclusive like mine or not, I think they are still useful* [R7].

One respondent also suggested a separate repository for data from the current IR and a linking service to enable one to link data to the published paper if it is available on the IR.

> *…we should separate the two, one platform for the repository and another platform for the research data, there can be linkages so that if you have the final output on the repository and you are interested in getting access to the research data a link can be provided at the button where you click on it and you will get access to the research data as well* [R4].

4.2.4.3. Staff development/Training

Another issue raised by some of the respondents was the need to build the capacity of support staff to support RDM. It is important to understand that most respondents believed

that staff generally had a fair idea about what RDM is, but this is not enough and therefore there is a need to provide RDM skills development opportunities for staff.

> *I believe there should be some formal training in that respect...Curation will have to come in, so I believe even if the person has some informal training,…there should be some formal training before the person is mandated to do that task* [R4].

> *I think we need to retool* [R3].

4.2.4.4. Staffing

Another issue that came up strongly in the data was inadequate staff, and as one interviewee indicated, there will be the need for dedicated staff to take up the new roles in RDM:

> *…you can't do everything looking at the limited resources in terms of human resources, we can't employ more because of the cap on employment…[*R5].

> *…so if we are going to do this then we'll need dedicated people, people whose job are just to go for these datasets and make sure they put them in an organised format and submit them on the platform* [R4].

4.2.4.5. Approach

Some of the respondents also suggest that evolving the culture of data publishing and sharing or institutional support for RDM should begin with internally funded research and those that the University is deeply involved.

> *…I think for the University, the first port of call should be the research that is actually funded by the University, funded with ORID money,…once they [researchers] do that researcher they should furnish ORID with the dataset and ORID should have a platform where they can keep these things* [R4].

> *Now that the University is collaborating with [Stellenbosch] looking at research uptake, I think we [should] look at research data security/ management especially, as I said, if for no data at all, those that the University is deeply involved. ISSER for example is a flagship of the University, Noguchi is a flagship, WACCI [West African Centre for Crop Improvement], I know the University is seriously involved in the contract and issues surrounding it so why not?, Because one day someone comes and say we are auditing and it can happen, and then they say you don't have any information* [R7].

4.2.4.6. Responsibility

Quite a number of stakeholders came up in the data gathered, especially the important role of top management in consenting and committing to supporting RDM. One respondent clearly expressed this position:

> *…it's not us sitting in UGCS making that call, the senior management, Vice-Chancellor, Pro-VC, the Deans, and head of departments must all agree that if this is what we are going to do then the right resources are put in there* [R1].

Another respondent indicated that the library is already playing roles that place it in a strategic position to contribute to the development of RDM at UG.

> *Of course, we have a role to play, so long as we are doing IR and creating metadata we can take this on also…when I think about the best place to host this. I think it will have to be the library because the library serves everybody and everybody when they want anything it is the library they come to* [R3].

The need for service providers to collaborate to deliver RDM support was also stressed.

> *I think working together with IT we can get a good storage for the data. This is where we need teamwork* [R3].

Responses on who should be spearheading RDM development was, however, not very clear even though the name of the library came up.

> *….we need to have a system in place, assigned to an office or a unit to do that and then we take it up from there* [R5].

> *I think someone needs to take it up, and it's ok for the library to spearhead it, so long us the librarians are educated on this as one of their roles in order to take it up, or a position should be created specifically for that or it could be added to the responsibility of the librarians in charge for IR. I don't think it will be bad if that responsibility were added on [to] the one responsible for the IR* [R3].

4.2.4.7. Sustainability

Another aspect that came up in the data analysis has to do the sustainability of supporting RDM at UG. The results show that most respondents believed it should be sustainable for the University to support RDM.

> *It should be sustainable because …elsewhere they are doing it and they have been doing it over the years... It's a matter of making it a priority and putting the right structure in place and that should be able to happen* [R1].

> *It should be feasible, especially these days that the University itself is spearheading grants, I don't see why not* [R7].

For one researcher, there had been a precedent, and so that provides useful lessons for the University.

> *…so [at Noguchi] we did things the University could not afford in those days, once we had our IGF [Internally generated Fund] and we could afford the servers, [and] you know our internet was the best until the recent developments, [so we developed those systems]…between 1996 and 1998 things were working there* [R7].

Another researcher, however, pointed to the fact that depending on external funding sources to develop RDM may be problematic.

> *As for that one, it's just the University that must decide this is what I must do, but in a third world country where we depend on people to give us what we want that is the problem*...[R6].

4.3. Discussion

The issues that emerge from the data analysis as presented in the previous section are discussed under this section. The discussions are based on the objectives of the study and organised under the following themes (Awareness and institutional response; Current activities and practices; Institutional capabilities; Requirements and other issues; Benchmarking for gaps in institutional awareness, support and capability; Priority areas for RDM development; Proposed Implementation plan). The discussions also attempt to draw parallels between the findings and the literature where possible.

4.3.1. Awareness and institutional response

Awareness is crucial in RDM development. This is well documented in the literature, particularly, the increasingly important role of librarians in promoting RDM awareness to facilitate uptake (Cox & Pinfield, 2014). The level of understanding about RDM and its significance is likely to affect stakeholder response to its development (Higman & Pinfield, 2015:372). Given that RDM is relatively new in this part of the world and that, as at the time of writing this dissertation there has been no study on RDM development in Ghana, participants were typically briefed on the concept mostly with a working definition of RDM. The results expectedly show a vague understanding about what constitutes RDM among respondents. At best, the term was used as synonymous with "data storage" and this is reflected in the examples of RDM specific supports that were provided by some respondents, especially, service providers. Other aspects (section 4.2.2.1) found to be germane to RDM were educed from the general research support that was offered. One respondent categorically admitted to the limited understanding of RDM among librarians.

> *I don't think we've understood it…there's some confusion…maybe there it's a slight understanding but I don't think we have understood it totally* [R3].
>
> *We'll need to know what data management is in the first place and we need to also understand why data management?* [R3].

These revelations, however, are not unusual. In the implementation of the "REsearch Data management for Mechanical Engineering Departments (REDm-MED)" project at the University of Bath in the UK, most researchers showed limited understanding of RDM and as in the case of UG often equated RDM to "research data storage" (Darlington et al., 2012: 8).

It is, however, worth highlighting the fact that respondents generally demonstrated some level of appreciation for the need for RDM. The data shows that respondents raised a number of critical issues like ramification for security, integrity, continuity and institutional reputation and this could serve as internal drivers for RDM development. Such responsiveness is good and desirable. What is even encouraging is the recognition of such ramifications at the institutional level as captured in its "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" (see section 4.2.1).

Of particular importance is the general recognition of poor data management as a risk. At least, for the two senior researchers who participated in this study, the risk of losing their data which one of them described as "a trade secret", was palpable. Data collection or creation can be expensive, time-consuming and arduous and in some scenarios involving observational measurements, it is practically impossible to recreate data. It is, therefore, natural for data lost to emerge strongly among the risks identified by respondents. It is also worth noting that some respondents also mentioned reputational risks to the institution where data cannot be made available for audit and verification, project failure as well as financial implications (see section 4.2.1.1). All these concerns are consistent with the literature (Harris-Pierce & Liu, 2012:599; Whyte & Tedds, 2011:2).

As hinted early on, the University's response to these issues is worth recognising. RDM is officially acknowledged as part of good research practice and is captured under section 5.6 of the "UG Research Policy" as a key statement and re-echoed in its "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management"; these will be discussed in greater detail in the succeeding section.

The data also indicates that management is fully interested in RDM. Despite the lack of institutional framework and systems for RDM, the keen interest shown by both the outgoing and incoming Vice-Chancellors towards the development of the HPC cluster at the department of Physics, together with the recognition of RDM as a key research policy statement, and the development of guidelines for good practices, are indicative of this interest. The data shows that the inability to develop appropriate systems for RDM as stipulated in the "UG Research Policy" can be attributed to human resource limitations, and

institutional priorities and exigencies. This revelation merits some critical examination, particularly, it would be interesting to investigate what issues shape UG's priorities for research development.

The University through ORID has developed a number of systems to facilitate research and research management at UG. The setting up of the Technology and Development Transfer Centre (TDTC) to facilitate issues of intellectual property and technology transfer, Committees for Ethical clearance to oversee and regulate the ethical conduct of research involving human subjects across the University as well as the development of concomitant policies and standard operating procedure are examples of such systems. These are fundamentally borne out of the main policy principles outlined in the "UG Research Policy". It is, therefore, understandable as one respondent expressed confidence that developing RDM systems could be next on the agenda and a natural extension [R5]. The development of such systems promotes good research practices and research integrity within the local research community.

These developments notwithstanding, a number of issues are still outstanding. It is not exactly clear among the researchers, at least for the two researchers who participated in the study, how the University is creating a commensurate environment to address the risks identified, as they clearly deny awareness of any internal systems or support to help avert or mitigate such risks:

> *…currently, I have no knowledge of any kind of systems like that to help curtail such risk* [R7].

While this may not necessary represent a true picture because of the scale of this study and the number of researchers engaged on the issue, it is likely that the University may have to step up its communication efforts and awareness drive within the local research community.

Furthermore, the data was silent on any existing institutional strategy for RDM development. Also, since writing DMPs is currently not an integral part of the research process at UG, it will be instructive to examine in greater depth than this study affords what the practices of the local research community are with regards to RDM and what the University's' aspirations and plan are for RDM development.

4.3.1.1. Institutional Policies and guidelines

Policies which specifically spell out key principles, expectations and roles for specific data management activities are fundamental to good RDM (Pryor, 2014:16-17; Rans & Jones, 2013:1). The institutional policy framework will normally include guidelines but these are often dynamic non-mandatory controls which typically evolve as services develop (Searle et al., 2015:444). The data shows that there is currently no specific policy on RDM at UG as there is for other research integrity issues like Ethics and Intellectual property which are also

key policy principles in the "UG Research Policy". The research policy was ratified by the University Council in November 2012 and addresses a few aspects of RDM. The "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management", which was approved by the ORID Management Board in 2013, however, spells out in greater details a number of best practices for guidance in the management of research data for the research community. However, these are not necessarily binding as one respondent suggested [R5] and do not represent a "public statement of intent and an expression of the commitment of management" (Pryor, 2014:16-17). Aspects of RDM are also loosely captured in the "UG Institutional Repository Policy", "UG Research Ethics Policy" and "UG Intellectual Property Policy".

Moreover, the data shows that the primary method for creating awareness about these policies and guidelines is through the ORID website. Links are also sent across the University community on its emailing platform and going forward they are considering compiling these into booklets and circulating them across the University. Despite these efforts, the two researchers indicated that they had no knowledge of any specific RDM policy or guideline. They both admit that it could exist but they were unaware. Indeed, the data confirmed no specific RDM policy is in place but the lack of awareness about an existing guideline for data management is remarkable. Again, while the scale of this study and the number of researchers engaged does not provide a very strong basis to presume this is the general picture, this revelation is still instructive and suggests that the University may have to step up its communication and awareness drive. The current approach to publicising these policies and guidelines, which is online, is logical given the current digital dispensation and the ease of reaching a much broader audience at a click: yet, the revelation of this study points to the contrary. This merits some consideration especially when such obliviousness is expressed by senior researchers with an average of two decades of research experience. Perhaps, interpersonal (face-to-face) level advocacy or College level engagements may help increase awareness.

Finally, the data does not reveal a clear institutional framework for monitoring compliance with these policies and guidelines provisions on RDM. As far as internal systems are concerned, satisfying the requirements for internal or external funding application seems to be how researchers' appreciation of, or adherence to, the policies, guidelines, and regulations are gauged. This may not be effective for monitoring compliance with RDM regulations and expectations.

4.3.2. Current activities and practices

Understanding prevailing practices and activities can be useful especially for streamlining and realigning current support in planning and developing a coherent RDM infrastructure and service. In order to get a sense of current activities and practices regarding RDM, the respondents from the service units were asked to describe how they support the research

process, and then usually a follow-up question about what specific RDM services were being offered. The researchers were also asked to describe how they manage their research data. A number of issues emerged.

First, respondents expressed varied opinions on current RDM services and support. While one respondent was convinced that RDM support was embedded in how the UGCS supported researchers, respondents from the Library said RDM support was not yet on offer. Neither researcher thought there were specific RDM services currently on offer through any of the service units, even though one did recount some efforts at one of the research centres of the University (Noguchi) several years ago to provide a central infrastructure for storing and protecting research data. Such variations could be attributed to misconceptions about what RDM is and what constitutes RDM support, and particularly, for researchers, a lack of awareness. Indeed, the data reveals a tenuous grasp of what RDM is among respondents and this often materialised in the kind of services frequently spoken about. As pointed out already, the existence of other services was deduced from the data on general research support.

To further advance the discussion on awareness about existing services and support for data management, it is important to understand that where existing support is offered at the project/research group level or even departmental level, it can impact the level of awareness of current services across the institution. This situation is quite different from the previous revelation about researchers' unawareness about existing data management guideline or policies relating to RDM, as such documents are relevant controls for ensuring consistency in institutional practices.

Finally, the result does not point to any formalized institutional services for supporting RDM, but they reveal some activities and practices that represent potentials that can be harnessed.

4.3.2.1. Institutional support

The data analysis reveals some institutional support for collaborative and computational research at UG, especially from the UGCS. In particular, the data shows that the University through UGCS provides huge storage facilities and high-speed network and processor base systems to facilitate such enterprise. The huge storage capacity of its HP Cloud infrastructure and the high-speed networking and processing capability of its HPC cluster was particularly emphasised by the respondent from the UGCS. These systems have obvious potential for research data storage, transfer, processing, and sharing.

More importantly, the examples cited to indicate such support suggest that these services are provided at the departmental or project level and explain why such support will not necessary be known to participants who may not be affiliated to that department or project team utilising such facilities. Even though the current HPC infrastructure is relatively small

and bespoke, it is also worth noting, the determination and commitment of the management of the University to grow, scale and optimise the current infrastructure to be harnessed by other academic units within the University. As the data shows, there is also currently effort at the department of Biochemistry, Cell and Molecular Biology under the College of Basic and Applied Science to establish an HPC cluster to run genomes and to provide other research computing services. Initially, the researcher considered including as one of the two researcher-participants, the Head of Department for this department [Biochemistry, Cell and Molecular Biology], who was also on the list of researchers provided by ORID to have met the criteria for selection. This potential participant was not immediately available because he was on an official duty abroad. Later, on his return, the researcher was invited to conduct the interview, but responses had already been gathered from other senior researchers who met the criteria and given the time constraint, the researcher could not conduct that interview. His contribution would have undoubtedly brought some nuance into the discussion particularly in relation to researcher's perception of existing institutional support for data management.

It is also refreshing to learn that as far back as two decades ago, the idea of a centrally managed infrastructure for research data storage, backup and sharing was being explored at one of the research centres of the university (Noguchi Memorial Institute for Research) (see section 4.2.2.1.4). Unfortunately, the current state of that facility could not be ascertained, but one researcher did comment on the difficulty the service faced due to a lack of an institutional framework. Such findings provide useful precedent and an opportunity for further exploration.

Even though there is no support for data deposit into the University's institutional repository, the data shows that it is being considered. The "UG Institutional Repository Policy" lists datasets as one of the materials permitted to be deposited, however, this is yet to be implemented. The main hindrance to this, as the data shows, is the lack of adequate personnel and with relevant curation skills. This appears to affect the drive to take it up. What was certain is that retooling existing staff involved in the IR service by providing them with training on RDM and data curation could be a useful starter for building capacity to support data storage and publishing using the IR.

The results also show support for data analysis. Apart from the HPC infrastructure which provides large-scale resources for high-speed data processing and analysis, tools and applications such as NVivo and SPSS for qualitative and quantitative data analysis respectively are available through the library and UGCS. These tools were, however, mentioned by only two participants from the Library and UGCS but were not corroborated by any of the researchers who participated. It would be useful to assess the awareness and attitude of faculty and research students regarding the utilisation of these facilities.

Finally, the results show that the University through ORID provides guidance on data management and advice on grant applications. Other aspects such as ethical research

conduct and Intellectual property issues are also dealt with. These issues are particularly pertinent for RDM service planning and development. ORID, among other things, alerts faculty members on funding opportunities and provides information about the requirements for such funding. It also, through the guideline for data management, provides specific information covering several aspects of RDM, such as data ownership, storage, documentation, retention, disposal, protection, and responsibility for data. There are, however, visible limitations that merit consideration. The fluid nature of the guideline leaves room for inconsistent practices in some key areas such as storage and retention and the guideline recognises the need for a clear retention policy.

> *"Nevertheless, it is important to have a clear retention policy that balances the best interests of society with those of the University and the individual researcher"*

Aspects like data management planning and locating or cataloguing existing institutional research data are not covered. The only data-related training that emerged from the result was SPSS training offered at the UGCS. However, the author workshops could be a good platform to introduce researchers to the data publishing requirement of some journal publishers.

4.3.2.2. RDM practices

Understanding researchers' everyday practices can be critical in aligning institutional RDM policies for effective adoption (Procter, Halfpenny & Voss, 2012:139-140). The scale of this study does not provide a detailed assessment of researcher practices at UG, but the findings still provide some general pointers which are worth considering.

It is important to note that there were divergent but also common grounds on issues relating to researchers' approach and attitude to research data and institutional RDM support, and this is significant because the researcher-participants represented two different disciplinary backgrounds (Botany and Pharmacy). Some of the key issues that emerged from the data on researchers' RDM practices are that researchers value their data enormously and find it absolutely important to keep it safe, but their ultimate goal is on publishing from it and not necessarily data management or sharing. This is understandable, because researcher's career progression and rewards are often tied to the research they conduct and the impact of their outcomes rather than on what they do with the data itself. One researcher-participant showed reluctance regarding sharing or publishing of data especially when he is yet to publish papers out it and so found institutional support for managing active data a bit uncomfortable. Another researcher-participant, however, supported the idea, noting the need for the University to have some control over the data generated by the research it underwrites. This particular researcher acknowledged that his opinion may have been influenced by his experience of working as an administrator within University at some point. Trust was another critical factor influencing researchers'

willingness to use central storage facilities and sharing data. The results also show that researchers will consider using internal and central infrastructure for managing their research data if they are confident of its effectiveness and guarantees for the safety of their data and if they retain greater control over decisions and choices of who to share it with. But it is clear that giving exclusive control over data management to researchers in the case of institutionally funded research data has ramifications for research continuity and data security. It is always important to ask oneself whether the people tasked with the responsibility for managing institutional data assets have the necessary skills and tools to effectively perform such tasks. If the answer is in the negative, then the university is exposed to the risk of losing valuable institutional assets.

These notwithstanding, the findings to a large extent are consistent with the literature. The respondent [R6] who showed reluctance to accept data sharing and using central storage systems because of mistrust is a senior researcher in the department of Botany which is a branch of the life sciences. In a study co-sponsored by the RIN and the British Library to "improve understanding of information use in the life sciences", it was found that life science researchers are generally reluctant to give away their data and may only share active data ("experimental data") on grounds that they are allowed sufficient time to publish their research from the data they generated and be given control in terms of personally publishing the data or making sharing choices rather that someone else doing it on their behalf (Williams et al., 2009:38-39).

As most researchers have not had the need to write data management plans, how will they respond to embedding DMP into the research process at UG? This will be interesting to investigate for the purposes of planning and developing services in the future.

The service providers generally see policy as a panacea to cultural change among researchers. They generally feel that a policy mandating researchers to deposit or share their research data is the only way to compel researcher to do that. Policy may be useful but the literature suggests that the policy should not be seen as a cohesive tool. Rather, incentives for sharing and depositing may be explored or internal systems may be developed to align with researchers' current practices (Procter, Halfpenny & Voss, 2012:139-140). This can be achieved where there are broad base consultations and engagements with the research community during the drafting or formulating stages of the policy.

Finally, looking at data practices from the ethical perspective also brings out some interesting issues. The only respondent who gave his opinion on the subject (as a service provider) said that the focus of the ethical systems was to protect the subjects from any potential injury posed by a study and not on data management. This is indeed interesting because the implication of understating the ethics of one's research is knowing how the data will be collected and handled to prevent any such injury. Such consideration for research data at such early stages of research is in itself a matter of RDM. Again, the

University recognise that there may be a need for institutional support in protecting data, and therefore admonishes researchers through the guidelines on data management to seek the approval of the University for their study to ensure such support is offered where necessary. The finding thus goes to buttress the point about the misconception and limited understanding of what constitutes RDM, as discussed early on. It also shows the University's interest in safeguarding institutional research data. However, there is little information on how the university is actualising this interest across the University.

4.3.3. Institutional capabilities

The adequacy of both human and technical capabilities, as well as financial resource availability, represent necessary institutional capabilities that input into the development of any successful RDM service (Whyte et al., 2014:285). Careful planning and sustained commitment from management are also required in the deployment of these resources. From the results, there are mixed feelings about the capability of the University to support RDM even though there is a general consensus that the University should be supporting RDM. Participants believed that support for RDM could be done sustainably but will require necessary retooling, putting relevant policies in place and institutional commitment. The results show that personnel do not have the necessary skills or adequate knowledge to fully support RDM and IT facilities are not necessarily provisioned for RDM purposes. It was beyond the remit of this study to conduct a data audit to determine the volumes of data generated within the UG research community or how much of it was known and available, but respondents generally feel that the existing storage capacities will suffice. It is not clear what the basis for such confidence is, but one respondent from the UGCS, as expected, showed an understanding of the situation, pointing out that a growing demand for research data storage and preservation will have implications for capacity to store them and so there will be a need for policy to regulate storage and disposal.

> *…if we start giving them storage we must have [a] policy on how long we are keeping the digital data, otherwise, within one year we would have generated so much in terabytes…for now our infrastructure apart from the network, is capable of handling [growing demand for storage]…it's a matter of having the right policies so…that when students generate data [and] they leave, you clear it and make it available for another fresh batch of students and that we will recycle it and provide that facility for students and staff* [R1].

What is more, the results do not show any particular understanding of the costs involved in RDM, yet, most respondents were confident that once the University makes RDM development a priority, it will be able to find the needed funding to support it just like it does for all other systems. However, the study could not establish any planned or real funding stream for RDM development at UG. One researcher-participant expressed worry about the difficulty in supporting such initiatives in third world countries without external

help and laments about the inadequacy of institutional funding opportunities for research. These are genuine concerns which have implications for RDM; first, RDM costs are mostly embedded in the project cost, so inadequate research funding can hamper RDM activities on the part of researchers, and second, when funding for initial development runs out, the University may need to internally fund such projects sustainably, thus, the need for an internal strategy for funding such a programme.

In all, it is safe to say that existing capabilities in relation to RDM support are crude, inchoate, uncoordinated and not formally instituted.

4.3.3.1. RDM knowledge and skills among service providers

The results show a contradictory response from the perspective of service providers. While some assume they possess some blend of skills that may be germane to RDM, others, particularly from the library are emphatic about the lack of capacity to support RDM especially in the areas of consultancy, training and curation. It will be better to consider these opinions in separate contexts rather than as a blanket depiction. In the library, it may appear somewhat exaggerated to suggest that there are no skills at all that can be utilised for RDM. The data showed that the library manages the institutional repository service, provides data analysis tools and also instruction; these are relevant to RDM though not enough to operate effectively. It could be that the tenuous understanding of the RDM concept among respondents explains why respondents in the library will presume they have no relevant skills. In fact, one respondent's initial reaction to the question of specific RDM support in the library was that the question should be directed to the people at IT.

The former appears to be a more realistic picture, as current skills from systems providing guidance in ethics and intellectual property (in the case of the research office) as well as support for data analysis, systems development, and storage (in the case of IT) are important aspects that can be included when considering RDM planning and development. Moreover, these views were likely shaped by the extent of data management responsibility these units have been exposed to. The research office has developed a guideline on RDM and a research policy capturing aspects of it; IT had to develop systems and provide storage for research groups who most likely will be storing their data on such systems, but the provision of analysis tools by the library is still not enough especially because RDM was mainly considered as storage of data.

What is certain, however, from all these opinions is that skills and knowledge of RDM is not adequate to support RDM: training is therefore required for support staff. The deficiency also creates new opportunities for curricula development. It may be important for the library school, in particular, to update its graduate curricula to reflect the emerging role of libraries as data scientists and data curators in this era of "the fourth paradigm".

4.3.3.2. IT infrastructure

According to Beitz et al. (2014:174) technical infrastructure for RDM can enable researchers to generate new knowledge and achieve a greater reach for their work, preserve and protect their research data, and facilitate collaborative research, data reuse and outcome validation. The study shows that UG has invested and continues to improve institution-wide technology capability to support different aspects of institutional process, activities and services. Of particular relevance to the purpose of this study, a number of systems and tools were identified that can be harnessed and extended to support data management. These systems and tools range from an HPC facility, a private cloud facility (HP Cloud Matrix), an institutional repository developed with DSpace software, and a robust network and security infrastructure and specific data analysis applications. It is important to establish that these systems are not necessarily provisioned for RDM, but represent potential for storage and preservation, analysis, access management, publishing and sharing.

The University has the capacity to develop specialised systems to facilitate collaborative research and data management. The UGCS has demonstrated this capability in the support it provides to some specific departmental or research group projects. A point in case is the development of an HPC facility to facilitate collaborative research at the Departments of Physics and African Studies. Such facilities provide complex processing and high-speed data transfer capabilities, as well as massive storage which can facilitate large-scale computational research.

There is a huge combined storage capacity within the University (about 150 TB), but how adequate this is to provide continuous support for research projects generating ever-growing volumes of data is not yet known. The respondent from the UGCS who provided insight into the technological capabilities appeared confident about the resilience of the current infrastructure, but two main issues are unresolved: first, the availability of adequate funding to acquire extra storage when necessary, and second, awareness and predictions about the current and future volumes of research data. It may also be necessary to view this in a broader institutional context since research data are not the only digital materials to be stored on these storage systems. Currently, the HP Cloud Matrix provides massive, high-speed networked storage capability of about 120TB and can be harnessed for active data storage and long term archiving.

The hosting of students' data on the Google platform (through an agreement between the University and Google) allows them access to free cloud service and storage space, to store, back up and sync their research data over the platform if they need to. This has also freed up space and resources to focus on staff. There is also an institutional backup service, but it appears to be provided for specific platforms under the control of the UGCS.

The IR currently supports publication deposit, but not research data even though it has the capacity to do so. The reason for this has been elucidated elsewhere but, to reiterate, lack

of adequate skills and human resources as well as institutional drive seem to be delaying the implementation of data deposit. The repository is developed on the DSpace software and according to the data, it has the capability to harvest and automatically ingest metadata from specific platforms. There are also specific data analysis tools such as Statistical Package for the Social Science (SPSS), for quantitative data analysis and NVivo, for qualitative data analysis available for researchers.

Finally, while there is currently no specific RDM platform developed at UG, systems developed by UGCS to support research projects, such as storage systems to safeguard research material including data, are designed to control and restrict access to only authorised people. One respondent justifies the effort of the UGCS in this regard:

> *…if people spend time and knowledge and brain power to develop anything and keep it within the systems, we must find a way of protecting it and make sure only the relevant people who have rights should have access to it* [R1].

4.3.4. Requirements and other issues

A common sense of agreement was expressed among respondents regarding the importance of RDM and the need for the University to support it and this was described as highly necessary. As a result, a number of areas came up strongly in the data which represent pressure points that need consideration. Some respondents recommended that RDM development and requirements to deposit or share research data should start with internally funded research and those in which the University has a high stake. Such recommendation is reflective of the shared appreciation of research data as an institutional asset among the participants.

The most advocated pressure point was policy development. Respondents saw policy as a key instrument that can give force to RDM systems and clarify the institution's position and expectations on RDM as well as drive compliance. Specifically, for aspects such as data deposit, retention, disposal and sharing, a policy was considered necessary for clarification and direction. As stated early on, apart from the guideline on RDM, there is no specific RDM policy and most of these aspects are not captured in the research policy. The only statement on retention in the research policy is that data must be kept for a minimum of ten years. However, there is no indication of what institutional framework is in place to ensure that this policy directive is adhered to.

Another pressure point identified in the data was the requirement for storage provision for institution research projects. This could be provided as a central service or departmental-level service. One respondent also suggested that a data repository is developed separately from the IR for dataset storage and linking facilities provided to link the research data to the published articles in the IR. More importantly, researchers are more likely to utilise such

facilities when they have greater control and are guaranteed about the security and privacy of their data and research materials.

Respondents often spoke about the need for a dedicated staff to handle data management issues, but one respondent suggested it could be added on to the portfolio of the IR librarian. What is most important is that the people tasked with the responsibility to support data management have the capability to do so and so capacity building is essential and inevitable. Many of the respondents have called for the need to provide RDM skill development opportunities for support staff to be able to fully support its development and operation at UG. Capacity building initiatives will have to clarify the concept of RDM and the full range of activities that it constitutes as well as the justifications and drivers of RDM development among HEIs. Particularly, the library may have to develop capacity to support in areas of consultancy, training and data curation, IT may have to look at the technical requirements of an RDM infrastructure and platforms, while the research office may also be exposed to the policy landscape and institutional frameworks for RDM.

The results also reveal an understanding among respondents that RDM development is a multi-stakeholder venture. The critical role of the University management came up strongly and this is to be expected because in all cases, management provides strategic direction and approves resources for service delivery. It is, therefore, natural to expect that any initiative that does not align with institutional strategy and priorities and thus lacks management's support will not succeed. One respondent also emphasised the need for collaboration among service providers, such as between the Library and IT.

Finally, on the issue of sustainability, the results indicate that most people were of the opinion that the University can sustainably support RDM. It is important, however, to consider this response through a particular lens. For most respondents, as pointed out early on, RDM support essentially represents the provision of storage infrastructure like a repository or server space for data to be deposited and accessed by whoever needs it. These are interventions over which the University is already making some efforts so, at best, what is needed is scaling up and extending such support to the broader university community. Another reason given by one respondent for that response was that, there was precedent where one of the research institutions within the University had set up its own basic infrastructure using its IGF to safeguard and provide a central storage facility for research data. This implies that it may be feasible for each department or college to work out possibilities of using their own internal resources to develop basic and specific systems to facilitate RDM at the local level. Indeed, the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" encourages departments to: *"try to create storage facilities for keeping data"*.

4.3.5. Benchmarking for gaps in institutional awareness, support and capability

Benchmarking helps to identify gaps between current and expected institutional RDM support and capabilities. It also provides an opportunity to focus RDM development. The EPSRC policy framework was used to benchmark existing practices, services and capabilities at UG (see chapter 2 section 2.6). The structure of the benchmarking exercise follows the outline for the DCC CARDIO Matrix for compliance with EPSRC policy framework and the optimal desirable expectation or level of development (*Embedding*) is adopted as the standard for comparison.

Best practices on RDM support capabilities based on EPSRC framework	Current Situation	Gaps
RDM Policy & Strategy Development: "The institution has ratified policy/strategy, and can demonstrate researchers' awareness of the regulatory environment, data policy principles and expectations, and of appropriate use of exemptions that may justify withholding research data"	There is currently no specific RDM policy, but RDM is captured briefly under section 5.6 of the UG Research Policy as one of its key policy statement as well as loosely in other policies like the "UG Institutional Repository Policy", "UG Research Ethics Policy" and "UG Intellectual Property Policy". The Research Office (ORID) makes available its research policies to the UG community mainly through its website. Researchers' awareness and appreciation of these policies and regulatory environment are essentially measured by their compliance with requirements when applying for funding through ORID. There is no clear strategy in place for RDM. Research conducted at UG complies with relevant regulations such as ethics, copyright, intellectual property and Academic Freedom. ORID through the TDTC and various Committees for Ethics has been facilitating issues on intellectual property and technology transfer, as well as oversight and regulation of the ethical conduct of research involving human subjects across the University. The awareness of RDM and policy provisions on RDM at UG is currently low. It may be necessary for the University to step up its communication and awareness creation	▪ No clear RDM policy and strategy ▪ Low level of awareness on RDM and policy provisions on RDM at UG. ▪ No clear institutional framework for monitoring compliance with the policy and guideline provisions on RDM

	efforts.	
Governance of data access & reuse: "There is an embedded culture of support at institutional level to ensure that sensitive data, access requests, and IPR are correctly handled at all stages of research lifecycles. Procedures and guidelines are continually reviewed to follow changes in legislation and best practice"	UG has an RDM guideline which covers some issues on data access but not on re-use. The policies addressing aspects of RDM also do not touch on the issue. The RDM guideline suggests that efforts must be made to protect and ensure the privacy of data with restricted access. It encourages the need to protect data that contains personal and sensitive information. The guideline, however, does not indicate the need to specify the conditions or criteria for access and re-use of such data. Researchers are encouraged to submit their research work for University approval so that any data protection need may be identified and appropriate support provided. It is not clear how UG is managing IPR and licensing for institutional datasets specifically. The guideline states that data may be stored externally (such as in external repositories) but does not specify additional modalities such as keeping record of metadata within the institution.	▪ RDM considerations (including data sharing and data deposit) are not fully integrated into the ethical approval process. ▪ No data registry or catalogue of institutional dataset at UG ▪ No licensing options for research data ▪ Research data are generally not published for other researchers to use/re-use
Business planning & sustainability: "Data management is adequately resourced and embedded in business planning to ensure continued support for curation; costs are transparent, controlled and recovered through direct or indirect support"	There are uncertainties surrounding the sustainability, costing and funding of RDM at UG. It is unclear whether RDM costs are understood and what strategies UG has in place to support RDM development, but RDM funding is expected to be borne out of UG's internal operational budget. This notwithstanding, no planned or real funding stream for RDM development at UG is known. The adequacy of current capabilities and infrastructure are also unclear.	▪ It is not clear what costs are involved in RDM and how these costs will be recovered ▪ There is no clear stream of funding for RDM development and support at UG.
Data Management & Sharing Plans: "The institution continually reviews	UG is interested in open access and sharing of research data among its faculty, however, there is currently no guidance on data management planning at UG. Writing a DMP is currently not an integral of the research	▪ There is no clear mandate for published papers to include a statement

available support for researchers to plan data management to meet funders' requirements, and to ensure data access statements are provided in research publications"	process and it is not a requirement for internal funding application. There is also no indication that UG is planning to embed DMP into the research process at UG anytime soon both for faculty and research students, but the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" recognises that some publishers may require that data that underscore a published article should be deposited in some public database or repository. Whilst UG subscribes to open access to research data, it expects that every data sharing decision is made with the permission of the University.	describing procedure and conditions for accessing research data.
Managing change & RDM service design: "The institution can identify its publicly funded research data holdings; risks are managed and policies and services monitor the need for change in provision"	RDM is considered a good research practice at UG and the University is fully aware of the risks associated with poor data management. The University is working towards creating a commensurate environment to address these risks and foster good practices. The University recognises the need to develop systems such as a "meta-database of research materials/ data repository" to support RDM as stated in the "UG Research Policy" but this is yet to be achieved. There is also no RDM-specific platform currently existing at UG for which reason it is currently unable to fully and comprehensively identify existing institutional data assets, their volumes, and locations. There are no known data curation services within UG and as RMD is not formally instituted, responsibilities are partially defined in the Guideline for RDM. The University places the primary responsibility for RDM and data stewardship on the researcher who is the Principal Investigator (PI), whiles the institution takes the responsibility of developing systems to support RDM.	▪ The RDM needs of researchers from the different disciplinary backgrounds at UG are unknown ▪ There is currently no clear institutional framework for managing research data at UG. ▪ Responsibilities are not fully clarified and defined ▪ No RDM platform or meta-database currently developed
RDM skills training and consultancy:	Currently, support staff do not possess the necessary skills or adequate knowledge to fully support RDM at UG. There are a few related	▪ Lack of adequate skills and competence to

"The institution has developed skills across relevant services, and capabilities are embedded in doctoral training and staff development, to provide secure and quality assured data mgmt & curation"	skills but those are not enough and therefore capacity building opportunities for the various support units is required to improve their understanding and skills to support RDM. Furthermore, RDM is not formally developed or instituted at UG but there are traces of activities and support that can be further explored and it is currently not fully integrated into the research process at UG, both for faculty and research students. There is currently limited support in the areas of consultancy, training and curation in terms of RDM.	fully support a coordinated data management service at UG ▪ Limited support services
Access & Storage Management: "The institution has effective services and infrastructure capabilities to support changing requirements for storage & access to digital and non-digital research data in active use"	UG has some level of technical infrastructure and capabilities that can support data storage (example a private cloud and an institutional repository). Departments are also encouraged to create systems for storing research data as stated in the "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management". Some facilities are also provided at UG to store non-digital data (example a herbarium at the Department of Botany to keep samples) but researchers are also encouraged to digitise non-digital data where possible.	▪ IT facilities are not necessarily provisioned for RDM purposes ▪ Existing capabilities in relation to RDM support are inchoate, uncoordinated and not formally instituted
Data cataloguing & publishing: "The institution has effective technical support for linking data assets to related information, and to create and record metadata about them with identifiers that may be cited over the long-term"	There is a desire to create a meta database for data assets within UG as stated in the "UG Research Policy", however, this has so far not been achieved. Research data are generally not published for other researchers to use/re-use.	▪ No record of data assets at UG ▪ No data linking and citing support yet available ▪ Research data are generally not published
Digital preservation	The "UG Research Policy" requires all PIs to	▪ No data registry available

& continuity : "Systems and services are established for managing research data for long-term (10 year+) data integrity and access to required security levels and are continually reviewed to ensure compliance with requirements"	retain research data for a minimum of ten years after the end of a research project. But there is currently no support available at UG for long-term preservation of research data. There are, however, infrastructure that can be extended to support long-term preservation, access, and integrity of institutional research data (example a Cloud Storage infrastructure – HP Cloud Matrix and Institutional repository are potentials).	▪ Systems and services are not formally established to support long-terms preservation

Table 4.1 Gap analysis of current RDM practices, systems and capabilities at UG

4.3.6. Priority areas for RDM development

Based on the findings on requirements, current institutional systems and capabilities, and the gaps identified based on the benchmarking exercise (sections 4.3.1 – 4.3.5), the following broad areas are proposed to focus RDM development at UG:

4.3.6.1. Initiating cultural change

The study findings show that RDM is currently not formally embedded in the research process at UG. Formally embedding RDM at UG will therefore inevitably trigger a cultural change in many cases. The results further reveal that the success of an RDM programme at UG hinges on "buy-in" and commitment from University management as well as researchers' receptiveness. Leadership is needed to drive a coordinated agenda for this change and to secure management support from the onset. Particularly, to engage with stakeholders and to boost interest, spearhead effective strategy development, and making a compelling case for RDM investment.

4.3.6.2. Institutional policy framework for RDM

The study found that respondents saw the need for policy to clarify the institution's position and expectations and to enforce them. There is currently no specific RDM policy at UG though aspects of it are loosely captured in other policy documents. Also, the existing guideline on RDM falls short in its coverage and the extent of treatment of some key aspects of RDM. It also lacks the binding effect since it is a non-mandatory instrument. A coordinated policy framework is, therefore, needed to define the University's priorities, provide direction for RDM development as well as clarify responsibilities.

4.3.6.3. Technical infrastructure for RDM

The data analysis show that there are some existing IT infrastructure at UG that has the potential for active and long-term data storage, access and publishing as well as for complex computational research. This can be further explored or streamlined to support data management requirements at UG. There may also be the need to develop new and coordinated technical infrastructure to foster data preservation, protection, discovery, re-use, and to facilitate collaborative research.

4.3.6.4. RDM skills and knowledge development

The study shows that generally support staff lack adequate skills and competency to fully support RDM at UG. Though a few skills were identified as germane these do not prepare most of them to fully support researchers. Capacity development opportunities and programmes will have to be provided to these professional staff to be able to acquire the necessary know-how and confidence to take up RDM as a new role. Similarly, embedding RDM in the research culture at UG will mean that opportunities for training must be created for UG researchers (both faculty and students).

4.3.6.5. RDM support services at UG

The study, despite its scale, shows that there are real opportunities for providing research data service at UG. It may be necessary to explore these opportunities to initiate basic services at UG in the short term while working towards a coordinated and mature service.

4.3.6.6. Funding

Sustainability of the RDM programme is in part dependent on the availability of a consistent and adequate stream of funds earmarked for RDM. The study could not establish any planned or real funding source for RDM development, but it is expected that funding such a project will have to come out of the institution's operational budget. An internal funding strategy may need to be put in place to facilitate RDM development and support at UG.

4.3.7. Proposed Implementation plan

As discussed in Chapter 2 section 2.6.1, an RDM roadmap outlines institutional goals or expectations and key milestones to achieving those goals. In section 4.3.5, the benchmarking exercise revealed some areas where UG is lacking in developing an effective RDM infrastructure and service. These areas have been categorised broadly under six priority areas where RDM efforts should be focused at UG (see section 4.3.6). Based on these six priority areas, specific activities are proposed to achieve the stated expectations. Furthermore, consideration is given to the potential of existing technical infrastructure (see section 4.3.3, section 4.3.3.1, section 4.3.3.2), the gaps in institutional awareness, capacity

and capabilities (4.3.5) as well as necessities for cultural change as discussed in chapter 2 section 2.2.4 and section 2.3.2.1.

A freehand approach has been adopted allowing the University management to decide what timelines are feasible and who should play which role rather than prescribing responsibilities and timelines for each task. This was considered appropriate partly because the study is a relatively small-scale study which included a limited number of stakeholders. Secondly, it will be more realistic to define timelines when a steering group has been constituted and they have considered the scale of resources required for all these activities.

Priority areas for action	Initiatives	Specific action	Milestones
1. Initiating cultural change	1.1. Establish an RDM steering group to spearhead and coordinate the development of an RDM at UG.	• Constitute a team of stakeholders from across the University community including the UGLS, UGCS, ORID, Ethical Committees, University Archive, Legal department and researchers. The team should be led by a champion for RDM, preferably the Pro-VC, RID. • Specify terms of reference and clarify responsibilities	▪ A steering committee of interested stakeholders assembled to advocate and coordinate RDM development at UG
	1.2. Embed RDM into local ethical clearance process at UG	• Collaborate with existing ethical committees to understand the current systems and process • Explore what issue and aspects of RDM can be infused into the current system • Embed RDM issues and questions into local ethical clearance approval process at UG	▪ Template for ethical clearance application including pertinent RDM issues and questions. ▪ Demonstrable understanding of implications of RDM for ethical research within the UG research community ▪ Responding to RDM questions as a key requirement for ethics approval at UG
	1.3. Promote re-use of UG research	• Advocate for data management planning, data	▪ A growing number of new research

	data	publishing and reuse at UG • Encourage UG research students to write DMP to enable data discovery, reuse and comprehension	outcomes based on data mining and/or meta-analysis
	1.4. Embed data management planning into internal grant application	• Investigate RDM costing models for grant applications • Embed RDM issues and questions into internal research funding application process at UG • Develop templates and guides indicating what RDM information that needs to be included in the DMP	▪ Template for internal grant application including pertinent RDM issues and questions. ▪ Clarification of RDM issues like identifying the usefulness of research data beyond the present investigation or for other discipline is a key consideration for awarding grant.
2. Institutional policy framework for RDM	2.1. Develop a coordinated policy framework for RDM	• Consult broadly with all key players within the University to get the different disciplinary issues and concerns on RDM • Draft policy framework • Identify and clarify responsibilities of key stakeholders (e.g. Researchers, UGLS, UGCS, ORID, Legal office, Management etc.) • Get appropriate approval (approval for policy must be from UG council)	▪ A Specific RDM Policy approved by University Council
	2.2. Communicate and promote awareness and compliance with the policy framework across UG	• Continue to communicate and create awareness about current and new policies and guidance on RDM through existing approach (ORID website and email notification) • Look out for new	▪ A clear framework for monitoring compliance is established ▪ Continuous engagement with UG research community to

		opportunity to improve awareness about policy framework through workshops, interpersonal (face-to-face) level advocacy or College level engagements • Develop an institutional framework for monitoring compliance with these policies and guidelines	drive awareness about RDM policy and compliance framework
3. Technical infrastructure for RDM	3.1. Align current technology and storage infrastructure with data management requirements	• Explore current potential of existing facilities and utilise them where feasible ○ HPC ○ IR ○ HP Cloud Matrix • Identify researcher requirements for data processing, storage and backup and align current technologies to support them	▪ Existing infrastructure with potential for RDM are aligned, formally instituted and promoted for RDM purposes
	3.2. Work towards developing a coordinated RDM infrastructure	• Conduct a data audit to understand the volumes of institutional research data available and the locations (such audit will include data emanating from longitudinal or policy research) • Investigate infrastructure needed for RDM across the UG • Explore and identify solutions that will support management of both digital and non-digital research data • Integrate requirements and solutions into developing a coordinated RDM system that is fit for purpose, and especially for collaboration	▪ A coordinated RDM platform is developed ▪ A catalogue of institutional research data is created ▪ IT solutions and mechanism that facilitate effective and secure data sharing is developed

	3.3. Develop specific tools and systems to support and facilitate data deposit, access, linking, citation, storage and publication to mention a few	• Develop a data registry (metadatabase) at UG to keep stock of metadata of all institutional research data • Develop easy to use tools for depositing as well as for citing, linking, integrating and publishing (CLIP) research data at UG	▪ A registry of institutional research data assets is developed
4. RDM skills and knowledge development	4.1. Develop RDM skill and knowledge of support staff at UG	• Develop internal programmes aimed at training support Staff on RDM: ○ UGLS ○ UGCS ○ ORID • Promote skill development through free online training courses such as MANTRA • Promote opportunities for support staff to network and participate in COP, locally, nationally and internationally	▪ Capacity to support, and provide guidance and training on RDM for researchers at UG ▪ Capacity for data curation ▪ Opportunities for RDM skills development are provided or supported when it is acquired through staff personal effort
	4.2. Develop RDM skill and knowledge of researchers (faculty and graduate students) at UG	• Provide opportunities for practical and discipline-focused training for researchers • Promote skill development through free online training courses such as MANTRA • Collaborate with Deans/Heads of Departments of Academic Units and University of Ghana Graduate school to embed RDM skills development into graduate training at UG	▪ Roll out skills development programme for researchers across UG ▪ RDM training embedded in Graduate School curricula
5. RDM support services at UG	5.1. Formalise and promote existing RDM support and services at UG	• Congregate existing and scattered services and functions through a single window • Promote these services and encourage patronage by	▪ An RDM portal created to link scattered RDM services and functions ▪ Improved user

		providing linkage between researchers with specific RDM needs and professionals who have the know-how to provide assistance.	uptake
	5.2. Develop new and simple RDM services	• Continue to provide storage service for specific project teams • Develop an RDM web portal that provides information on best practices and provides links to institutional RDM policy framework, support as well as useful external resources and tools • Set up help desk systems that can be used to route RDM queries • provide dedicated email addresses or details of contact persons for RDM support	▪ An RDM portal created to provide simple visual guidance on research data creation, storage and management ▪ Services available for consultancy, training and data management planning support
6. **Funding**	6.1. Introduce RDM programme into institutional planning cycle at UG	• Make a compelling business case for RDM that is aligned with institutional strategy to University council • Develop an RDM business plan • Identify and include departmental requirements for RDM into institutional planning • Explore at the School/departmental level possibilities of using part of their IGF or operational budget to develop and support specific systems to facilitate RDM at the local level	▪ Draft business plan for RDM ▪ Local level initiatives and systems developed to support departmental RDM ▪ Departmental requirement for RDM is included in institutional planning
	6.2. Explore strategies for cost recovery and RDM financing	• Investigate RDM costs and funding models to support long-term preservation and curation of research data	▪ Clearly outlined cost variables for RDM costing in project grant application

			▪ Information on RDM costing included in the guidance for funding application provided by ORID ▪ Information about RDM costing and financing options made available through RDM web portal

Table 4.2 Proposed Implementation plan for RDM at UG

4.4. Conclusion

In this chapter, the data collected in the course of the study were analysed, presented and discussed. Several issues and aspects emerged from the analysis and these were discussed under seven main sub-headings: awareness and institutional response, current activities and practices, institutional capabilities, requirements and other issues, benchmarking for gaps, priority areas for RDM development and proposed Implementation plan. In the next and concluding chapter, the study will be summarised, concluded and recommendations made for consideration.

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1. Introduction

In the previous chapter, Chapter four *Analysis of findings and discussions*, the data collected were analysed and discussed. This concluding chapter summarises the salient findings and issues that emerged in the study, aligning them with the objectives of the research work. The study is concluded and based on the findings, recommendations are made for consideration by management in developing RDM at the University of Ghana.

5.2. Summary of findings

The aim of the study was to assess the state of the art of RDM development and institutional preparedness at UG in order to develop a roadmap for implementation at UG. Specifically, the study sought to achieve the following objectives:

1. To assess data management requirements, activities and capabilities at UG
2. To benchmark existing practices, systems and facilities against that of EPSRC policy framework to identify any gap in terms of institutional awareness, support, capacity and capability.
3. To identify priority areas for RDM development based on identified strengths and weaknesses
4. To propose an action plan for RDM implementation at UG

To achieve these objectives, a total of seven respondents were interviewed on existing policies, procedures and systems; technology infrastructure; knowledge, skills and capabilities; and support services relating to RDM at UG. Triangulation was used to test for consistency, corroborate evidence, and draw convergence in the data. The findings of the study are summarised below.

5.2.1. Data management requirements, activities and capabilities at UG

5.2.1.1. Requirements

The findings of the study show that respondents agree that RDM is critical to the research process and that the University should be actively involved, particularly in supporting its development here at UG. Despite this being a small-scale study, a few requirements emerged that were advocated by the respondents (both researchers and services providers).

5.2.1.1.1. Policy

The study revealed that developing a policy on RDM was considered key in clarifying UG's position and expectation on the subject as well as to encourage compliance within its research community. Though the issue of RDM was briefly captured in the "UG Research Policy", there are still key aspects such as on data deposit, retention, disposal and sharing that needs clarification and direction. The "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management" which addresses RDM to a greater extent is largely a non-mandatory instrument and is thus not binding and does not serve the full purpose of a policy.

5.2.1.1.2. Storage service

The findings of the study revealed that there is a need for some level of institutional data storage system to be provided whether at a central or departmental level to store and protect institutional research data. This will provide UG some control over the data generated through research that it funds. Particularly for active data storage, the analysis shows that there is the need for embargo and effective security infrastructure with respect to depositing and accessing research data on such central storage platforms. A data repository (separate from the existing IR) was also identified as needed for long-term storage of research data.

5.2.1.1.3. Staffing and RDM skill development for support staff

There is a need to employ more people, particularly at the Research Office and the Library to deal with aspects of RDM support. Over and above this requirement, the results indicates that there is a real need for RDM skill development opportunities for support staff (Research Office, IT Services and the Library) to be able to fully support RDM development and operation at UG.

5.2.1.2. Activities

RDM is not formally developed or instituted at UG but there are traces of activities and practices that are noticeable. The concept of RDM is vaguely understood but there is a fair appreciation of the risks associated with poor data management practices, particularly, the possibility of losing data and the reputational risk it poses to UG. A number of institutional RDM supports were identified as potentials for further exploration at UG. These include support for collaborative and computational research, provision of specific tools and support for data analysis, guidance on RDM (specifically on funding application and best practices on data management), as well as storage service for some research projects. There is also support in the area of Intellectual property and ethics but these are yet to be streamlined with RDM. In principle, the idea of an institutional support data management it not exactly new in the UG environment as semblance of same had been explored and tried

several years ago in one of its research centres (Noguchi Memorial Institute for Medical Research). The requirement for DMP is currently not an integral part of the research process at UG and it is not clear whether the University will be considering including DMPs as a requirement for institutional grant application in the near future. What is more, no institutional strategy for RDM development or implementation was identified.

5.2.1.3. Capabilities

The following capabilities have been identified at UG that can be harnessed for RDM development.

5.2.1.3.1. Policy

There is currently no specific RDM policy at UG but RDM is captured and treated briefly as one of the policy statements in the "UG Research Policy" which was ratified in November 2012 by the University Council. There is, however, a guideline on RDM which provides guidance on best practices but this is not binding per se on researchers. The current mode of communicating policies and guidelines on research and data management has been mainly through the ORID website with links being sent through the University emailing platform to notify the university and local research community about these policies and guidelines. However, it appears that this has not been sufficient and the University may need to step up its communication and awareness drive. There is also no clear institutional framework for monitoring compliance with these policies and guidelines and the current improvisations do not make for effective monitoring of regulations and expectation.

5.2.1.3.2. Technical and technological infrastructure

The findings show that there are some existing systems, tools and infrastructure that are not necessarily provisioned for RDM but can be harnessed and extended to support RDM at UG. These systems mainly enable the storage and back-up of active research data as well as long-term data storage and publishing. There are also tools and capabilities for small-scale, large-scale and complex data analysis, high-speed data transfer and massive storage which can facilitate large-scale computational research.

The following infrastructure were identified through the study:

- An HPC facility
- A private cloud facility (HP Cloud Matrix)
- An institutional repository (UGSPace) developed with DSpace software
- Institutional Google Drive platform (through collaboration with Google)
- SPSS
- NVivo
- A robust network and security infrastructure

5.2.1.3.3. Resourcing

The study could not establish any planned or real funding stream for RDM development at UG. It is generally expected that developing RDM will be absorbed and funded by the University much in the same way it has been doing for other systems that have been developed over the years.

5.2.1.3.4. Staffing, knowledge and Skills

The analysis of the data revealed that currently support staff do not possess the necessary skills or adequate knowledge to fully support RDM at UG. There were a few skills that could be harnessed but these are not enough and therefore capacity building opportunities for the various support units is required to improve their understanding and also acquire relevant skills to support RDM. While there have been calls to employ more and dedicated staff to handle RDM, it is clear that retooling existing support staff by providing specific RDM training could be a useful starter.

5.2.1.3.5. Sustainability

The study revealed a few service requirements particularly the provision of an institutional storage system and also the development of policy, but to get a better picture of the needs of the different disciplinary groups within the local research community, a full-scale data audit and requirement gathering exercise may have to be conducted at the project or departmental level across the University. The study does not reveal any particular understanding of the costs involved in RDM and it is unclear how exactly the University will be supporting it. While the results show that there is confidence in the resilience and robustness of the current infrastructure, it is still unclear whether existing storage capabilities are adequate for storing institutional research data assets. This is so because there is no clarity on the availability of adequate and sustainable funding stream to acquire extra storage when necessary, and also there is no clarity about awareness and predictions about the current and future volumes of research data.

5.2.2. Gap analysis

The benchmarking exercise revealed a number of gaps, these are outlined below:

- No clear RDM policy and strategy
- Low level of awareness on RDM and policy provisions on RDM at UG
- No clear institutional framework for monitoring compliance with the policy and guideline provisions on RDM
- RDM considerations (including data sharing and data deposit) are not fully integrated into the ethical approval process

- No data registry or catalogue of institutional dataset at UG
- No licensing options for research data
- Research data are generally not published for other researchers to use/re-use
- It is not clear what costs are involved in RDM and how this costs will be recovered
- There is no clear stream of funding for RDM development and support at UG
- There is no clear mandate for published papers to include a statement describing procedure and conditions for accessing research data
- The RDM needs of researchers from the different disciplinary backgrounds at UG are unknown
- There is currently no clear institutional framework for managing research data at UG
- Responsibilities are not fully clarified and defined
- No RDM platform or meta-database currently developed
- Lack of adequate skills and competence to fully support a coordinated data management service at UG
- Limited support service
- IT facilities are not necessarily provisioned for RDM purposes
- Existing capabilities in relation to RDM support are inchoate, uncoordinated and not formally instituted
- No data linking and citing support yet available.

5.2.3. Priority areas for RDM development

The following priority areas were identified to focus RDM effort at UG

- Initiating cultural change
- Institutional policy framework for RDM
- Technical infrastructure for RDM
- RDM skills and knowledge development
- RDM support services at UG
- Funding.

5.2.4. Implementation plan

The six broad areas, identified as critical to focus RDM efforts at UG, in section 4.3.6 were further cascaded into a number of specific initiatives and tasks for implementation. These were based on the potential of current infrastructure, gaps in institutional awareness and capability, and also essentials for cultural change. Key milestones were defined for each initiative. The study identified the following initiative for implementation:

- Establish an RDM steering group to spearhead and coordinate the development of an RDM at UG
- Embed RDM into local ethical clearance process at UG
- Promote re-use of UG research data
- Embed data management planning into internal grant application
- Develop a coordinated policy framework for RDM
- Communicate and promote awareness and compliance with the policy framework across UG
- Align current technology and storage infrastructure with data management requirements
- Work towards developing a coordinated RDM infrastructure
- Develop specific tools and systems to support and facilitate data deposit, access, linking, citation, storage and publication to mention a few
- Develop RDM skill and knowledge of support staff at UG
- Develop RDM skill and knowledge of researchers (faculty and graduate students) at UG
- Formalise and promote existing RDM support and services at UG
- Develop new and simple RDM services
- Introduce RDM programme into institutional planning cycle at UG
- Explore strategies for cost recovery and RDM financing.

5.3. Conclusion of the study

The overarching aim of the study is to assess the state of the art of RDM development and institutional preparedness at UG through existing activities and capabilities. Specifically, three questions were answered: whether there are existing and well-implemented policies

and procedures; whether there are adequate technologies; and whether resources (financial, human and skills) are sufficient to support effective RDM at UG. Answers to these questions provided a general understanding of the current situation at UG in terms of RDM.

The findings of the study revealed that RDM at UG is currently underdeveloped but with immense potential for growth. A number of RDM related activities were identified. There is no specific RDM policy but aspects of it are loosely covered in other policies, there is also an RDM guideline available. However, it appears UG may have to step up its efforts in driving awareness about existing policies and guidelines on RDM. A few technical infrastructure were also found to be germane to RDM but existing capabilities were generally found to be inchoate, uncoordinated and not formally instituted. In terms of resourcing, there was no clear stream of funding for RDM, and there was a lack of adequate skill and knowledge among professional staff to fully support RDM. The study thus shows that RDM is a natural part of HEI's research activities, even when they are not formally instituted.

These results are benchmarked against the EPSRC policy framework to identify gaps. Based on these gaps and other findings, the study proposed priority areas for consideration and developed an action plan for implementation.

5.4. Recommendations

Over and above the priority areas identified in chapter 4 section 4.3.6, the following recommendations are made for further development of RDM at the University of Ghana:

- Define a workable timeframe and assign responsibilities for key actions outlined in the implementation plan to specific stakeholders or units within the UG
- Start with the low hanging fruits. Explore short-term potentials and invest in infrastructure incrementally to reduce the duress of the huge initial capital outlay
- Active engagement with the academic community at UG throughout the RDM development process is essential to get them on board from the start and to ensure that eventual solutions are fit for purpose
- Forge strong partnerships with stakeholders and foster active interdepartmental communication and collaboration
- Policy development should involve a broad-based consultation and input from across the wide spectrum of the University community in order to increase acceptance and to ensure that the varying disciplinary issues and concerns are taken into consideration
- Explore incentives to drive RDM uptake among the different disciplinary groups at UG

- There are several templates and transferable toolkits available from the DCC for adoption and re-use. These should be explored to provide a quick and useful head start for UG
- As much as possible RDM development at UG should focus on delivering discipline-specific support and guidance
- Policies must be developed for data selection and data retention, this is important because not every data can be and should be retained, stored, managed and reused and criteria for making such decision should be clearly laid out in the policy
- Work towards establishing a data repository with the capability to link datasets with corresponding article on both internal and third party systems
- Going forward, UG may explore external funding opportunities to help support RDM systems development and services (particularly, institutions that advocate and support open access and research development initiatives)
- Explore external and national level infrastructure that can be used to support and supplement institutional effort towards RDM development, particularly, for collaborative research
- Explore DCC DMP Online and UC3 DMPTool for developing local requirement and questions for writing DMPs
- Investigate requirements for interoperability between existing internal systems and third-party service providers.

5.4.1. Suggestions for further research

The findings of this study have some implications for future studies. Therefore, the following suggestions are made for consideration in future and further research:

i. The study limited stakeholders to Researchers, IT services, Research Office and the Library. Further studies can expand to include other critical stakeholders like the University Executive management, legal department, archives, Government and so on

ii. The relatively small-scale of this study and the number of researchers engaged does not provide a strong picture of researchers' practices and requirement at UG. Future studies can look at expanding the number of researchers engaged and the assessment should be done at the departmental, research group or project level to get a more realistic picture across the institution

iii. Future research should explore opportunities at the national level and their implications for institutional RDM development and collaborative research.

5.5. Summary of chapter

This chapter provided a summary of the findings of the study, conclusions and recommendations. The study achieved its aim of assessing the state of the art of RDM development at the UG. It identified existing RDM activities and capabilities even though there were no formal RDM structures and systems. The study thus shows that RDM is a natural part of HEI's research activities, even when they are not formally instituted. The study also proposed an implementation plan for the UG which covers critical areas that require attention based on the requirements, activities and capabilities identified through empirical evidence. This is useful for the UG in focusing its efforts and limited resources in developing a coordinated and coherent infrastructure. The study contributes to the body of knowledge on institutional RDM implementation from the perspective of a developing nation, specifically, Ghana and provides a useful case study for other institutions who may want to develop an RDM programme.

REFERENCES

Akers, K.G. & Doty, J. 2013. Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, *8*(2): 5-26.

ANDS. n.d. *About ANDS*. [Online]. Available: http://www.ands.org.au/about-ands.html#ardc (Accessed: 22 January, 2016).

Arghode, V. 2012. Qualitative and quantitative research: paradigmatic differences. *Global Education Journal*, 2012(4):155-163.

Auckland, M. 2012. *Re-skilling for research: an investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers*. [Online]. Available: http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf (Accessed: 25 February, 2016).

Australian Government. 2007. *Australian code for the responsible conduct of research.* [Online]. Available: http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf (Accessed: 22 January, 2016).

Australian Government. n.d. *Research data service (RDS).* [Online]. Available: https://www.rds.edu.au/ (Accessed: 12 February, 2016).

Awre, C., Baxter, J., Clifford, B., Colclough, J., Cox, A., Dods, N., Drummond, P., Fox, Y. *et al*. 2015. Research data management as a "wicked problem". *Library Review*, *64*(4/5): 356-371.

Ball, J. 2013. Research data management for libraries: getting started. *Insights*, 26(3): 256-260.

Beagrie, N., Chruszcz, J. & Lavoie, B. 2008. *Keeping research data safe: a cost model and guidance for UK universities*. [Online]. Available: http://www.webarchive.org.uk/wayback/archive/20140615221657/http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf (Accessed: 22 January, 2016).

Beagrie, N., Lavoie, B. & Woollard, M. 2010. *Keeping research data safe 2: final report.* [Online]. Available: http://www.webarchive.org.uk/wayback/archive/20140615221405/http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf (Accessed: 22 January, 2016).

Beitz, A., Dharmawardena, K. & Searle, S. 2012. *Monash University research data management strategy and strategic plan 2012-2015*. [Online]. Available: https://confluence-vre.its.monash.edu.au/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-

publicrelease.pdf?version=1&modificationDate=1334289180000 (Accessed: 12 February, 2016).

Beitz, A., Groenewegen, D., Harboe-Ree, C., Macmillan, W. & Searle, S. 2014. Case study 3: Monash University, a strategic approach. In: Pryor G., Jones S. & Whyte A. (eds.), *Delivering research data management services: fundamentals of good practice.* London: Facet Publishing: 163-190.

Berman, F., Lavoie, B., Ayris, P., Cohen, E., Courant, P., Dirks, L. & Friedlander, A. 2010. *Sustainable economics for a digital planet: ensuring long-term access to digital information: final report of the Blue Ribbon Task Force on sustainable digital preservation and access*. [Online]. Available: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (Accessed: 19 January, 2016).

Bhattacherjee, A. 2012. *Social science research: principles, methods, and practices*. [Online]. Available: http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks (Accessed: 15 October, 2015).

BIS. 2010. *Delivering the UK's e-infrastructure for research and innovation.* Department of Business, Innovation and Skills. [Online]. Available: http://www.rcuk.ac.uk/RCUK-prod/assets/documents/research/esci/e-Infrastructurereviewreport.pdf (Accessed: 26 January, 2016).

BIS. 2011. *Report of the e-Infrastructure advisory group.* Department of Business, Innovation and Skills. [Online]. Available: http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/e-IAGreport.pdf (Accessed: 26 January, 2016).

Bishoff, C. & Johnston, L. 2015. Approaches to data sharing: an analysis of NSF data management plans from a large research university. *Journal of Librarianship and Scholarly Communication*, *3*(2):1-27.

Borgman, C.L. 2012. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6): 1059-1078.

British Library. 2016. Previous workshops. [Online]. Available: http://www.bl.uk/aboutus/stratpolprog/digi/datasets/workshoparchive/archive.html (Accessed: 19 February, 2016).

Brown, S., Bruce, R. & Kernohan, D. 2015. *Directions for research data management in UK universities*. UK: Jisc. [Online]. Available: http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf (Accessed: 10 February, 2016).

Carlson, J. 2014. The use of life cycle models in developing and supporting data services. In: Ray, J. M. (ed.) *Research Data Management: Practical Strategies for Information Professionals.* Series: Charleston Insights in Library, Information, and Archival Sciences. West Lafayette, IN: Purdue University Press: 63-86.

Charbonneau, D.H. 2013. Strategies for data management engagement. *Medical Reference Services Quarterly*, *32*(3):365-374.

Christensen, L. B., Johnson, R. B. and Turner, L. A. 2011. *Research methods, designs, and analysis*. 11th (Eds.). Boston: Allyn and Bacon, Pearson.

Christensen-Dalsgaard, B., van den Berg, M., Grim, R., Horstmann, W., Jansen, D., Pollard, T., & Roos, A. 2012. *Ten recommendations for libraries to get started with research data management: final report of the liber working group on e-science/research data management.* [Online]. Available: http://libereurope.eu/wp-content/uploads/WGSC_20120801.pdf (Accessed: 13 August, 2015).

Collins, E. 2012. The national data centres. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 151-172.

Collins, S., Beard, L., Besson, J., Finch, J., Goff, M., Halfpenny, P., Grahame, T., McDerby, M. et al. 2010. *Towards a generic research data management infrastructure*. In: Proceedings of the 9th UK e-Science All Hands Meeting (AHM 2010), Cardiff, 13-16 Sep 2010. [PDF]. [Online]. Available: http://wrap.warwick.ac.uk/52815/1/WRAP_Procter_Towards%20a%20generic%20infrastructure%20all%20hands%20meeting%202010.pdf (Accessed: 12 February, 2016).

Corrall, S. 2012. Roles and responsibilities: Libraries, librarians and data. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 105-134.

Corrall, S., Kennan, M.A. and Afzal, W. 2013. Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends*, *61*(3): 636-674.

Corti, L., Van den Eynden, V., Bishop, L. & Woollard, M. 2014. *Managing and sharing research data: A guide to good practice*. London: Sage Publication Ltd.

Cox, A. M. & Pinfield, S. 2014. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4): 299-316.

Cox, A. M., Pinfield, S. & Smith, J. 2016. Moving a brick building: UK libraries coping with research data management as a 'wicked' problem. *Journal of Librarianship and Information Science,* 48(1):3-17.

Cox, A., Verbaan, E. & Sen, B. 2012. Upskilling liaison librarians for research data management. *Ariadne*, (70).

Cragin, M. H., Heidorn, P. B., Palmer, C. L., & Smith, L. C. 2007. *An educational program on data curation*. A conference poster presented at the American Library Association Conference, Science and Technology Section, Washington, D.C., June 25, 2007.[PDF]. [Online]. Available: https://www.ideals.illinois.edu/handle/2142/3493 (Accessed: 12 January, 2016).

Creswell, J. W. 2009. *Research design: qualitative, quantitative, and mixed methods approaches*. 3rd ed. London: Sage.

Creswell, J. W. 2013. *Qualitative inquiry and research design: choosing among five approaches*. 3rd ed. Los Angeles, CA: Sage.

DAF. 2009. *Data Asset Framework: implementation guide*. [Online]. Available: http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf (Accessed: 21 February, 2016).

DAF. n.d. *Four steps to effective data management*. [Online]. Available: http://www.data-audit.eu/index.html (Accessed: 21 February, 2016).

Daniel, J., Sawyerr, A., Addy, M. E., Akyeampong, E., Ampiah, A. P., Archampong, E.Q., Desai, U. B., Falase, A.O. et al. 2007. *Report of the Visitation panel to the University Of Ghana*. [Online]. Available: http://oasis.col.org/bitstream/handle/11599/230/UofGhana.pdf?sequence=1&isAllowed=y (Accessed: 16 August, 2016).

Darlington, M., Ball, A., Thangarajah, U., McMahon, C. & Lyon, E. 2012. REDm-MED Project Final Report to JISC, (version 1.1). University of Bath. [Online]. Available: http://www.ukoln.ac.uk/projects/redm-med/reports/redm-med_final_report_v1.pdf (Accessed: 22 November, 2016).

Davidson, J., Jones, S., Molloy, L. & Kejser, U.B. 2014. Emerging good practice in managing research data and research information within UK universities. *Procedia Computer Science*, *33*:215-222.

DCC. 2011a. *Collaborative assessment of research data infrastructure and objectives (CARDIO)*. [Online]. Available: http://www.dcc.ac.uk/projects/cardio (Accessed: 23 February, 2016).

DCC. 2011b. *CARDIO tool*. [Online]. Available: http://cardio.dcc.ac.uk/ (Accessed: 23 February, 2016).

DCC. 2011c. *Collaborative assessment of research data infrastructures and objectives: workflow planning & documentation*. [Online]. Available: http://www.dcc.ac.uk/sites/default/files/documents/Cardio_Characteristics_By_Statement.pdf (Accessed: 23 February, 2016).

DCC. 2012. *DMP Online*. [Online]. Available: https://dmponline.dcc.ac.uk/ (Accessed: 12 February, 2016).

DCC. 2013a. *Roadmaps to EPSRC expectations on research data*. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps (Accessed: 11 February, 2016).

DCC. 2013b. *UK institutional data policies*. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies (Accessed: 11 February, 2016).

DCC. 2013c. *Policy tools and guidance*. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/policy-tools-and-guidance/policy-tools-and-guidance (Accessed: 11 February, 2016).

DCC. 2013d. *Checklist for a data management plan*. v.4.0. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/resources/data-management-plans/checklist (Accessed: 12 February, 2016).

DCC. 2013e. *CARDIO RDM matrix: assessing institutional awareness and readiness for compliance with EPSRC policy framework*. [Online]. Available: http://www.dcc.ac.uk/sites/default/files/documents/events/EPSRC-workshop/CARDIO_matrix_EPSRC.pdf (Accessed: 23 February, 2016).

DCC. 2016a. *Disciplinary RDM training*. [Online]. Available: http://www.dcc.ac.uk/training/train-trainer/disciplinary-rdm-training/disciplinary-rdm-training (Accessed: 12 February, 2016).

DCC. 2016b. *RDM for librarians*. [Online]. Available: http://www.dcc.ac.uk/training/rdm-librarians (Accessed: 12 February, 2016).

DCC. 2016c. *Data management and curation education and training*. [Online]. Available: http://www.dcc.ac.uk/training/data-management-courses-and-training (Accessed: 12 February, 2016).

DCC. n.d.a. *DCC curation lifecycle model*. [Online]. Available: http://www.dcc.ac.uk/resources/curation-lifecycle-model (Accessed: 1 February, 2016).

DCC. n.d.b. *What is digital curation?* [Online]. Available: http://www.dcc.ac.uk/digital-curation/what-digital-curation (Accessed: 12 January, 2016).

DCC. n.d.c. *RDM guidance webpages*. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages (Accessed: 13 February, 2016).

Delserone, L. M. 2008. At the watershed: preparing for research data management and stewardship at the University of Minnesota Libraries. *Library trends*, 57(2): 202-210.

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.

DIRISA. n.d. *Data intensive research initiative of South Africa (DIRISA).* [Online]. Available: http://www.dirisa.ac.za/ (Accessed: 12 February, 2016).

Donnelly, M. 2012. Data management plans and planning. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 83-104.

Donnelly, M. 2014. *Five steps to developing a research data policy: DCC 'Quickstart' leaflets*. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/five-steps-developing-research-data-policy/five-steps-developing-research (Accessed: 11 February, 2016).

DST. 2013. *National integrated cyberinfrastructure system: a framework for the establishment and maintenance of a sustainable NICIS*. A report to the Department of Science and Technology (DST), Republic of South Africa. [Online]. Available: http://www.dst.gov.za/images/NICIS_Framework_Report_December_2013_26_May_2014.pdf (Accessed: 12 February, 2016).

Eckard, M. 2014. *Best Practices for Research Data Management*. [Online]. Available: http://works.bepress.com/maxeckard/8/ (Accessed: 19 January, 2016).

Ekmekcioglu, Ç. and Rice, R. 2009. *Edinburgh data audit implementation project*. Scotland: University of Edinburgh. [Online]. Available: *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.175.6624&rep=rep1&type=pdf* (Accessed: 22 February, 2016).

EPSRC, 2011. *EPSRC policy framework on research data*. [Online]. Available: https://www.epsrc.ac.uk/about/standards/researchdata/ (Accessed: 23 January, 2016).

EPSRC. 2015. *Scope and benefits*. [Online]. Available: https://www.epsrc.ac.uk/about/standards/researchdata/scope/ (Accessed: 1 October, 2015).

Flick, U. 2011. *Introduction to research methodology: a beginner's guide to doing research project*. London: Sage.

Flick, U. 2014. *An Introduction to qualitative research*. 5th ed. Los Angeles, CA: Sage.

Fowler, S. 2012. *Managing research data: a pilot study in health and life sciences*. [Online]. Available: http://www1.uwe.ac.uk/library/usingthelibrary/researchers/manageyourresearchdata/managingresearchdata.aspx (Accessed: 24 February, 2016).

Friedlander, A. & Adler, P. 2006. *To stand the test of time: Long-term stewardship of digital data sets in science and engineering*. [Online]. Available: http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf (Accessed: 23 January, 2016).

Gelo, O., Braakmann, D., & Benetka, G. 2008. Quantitative and qualitative research: Beyond the debate. *Integrative Psychological and Behavioral Science*, 42(3): 266-290.

Gerring, J. 2007. *Case study research: principle and practices*. Cambridge: Cambridge University Press.

Gibbs, H. 2009. *Southampton data survey: our experience and lessons learned*. JISC. [Online]. Available: http://www.disc-uk.org/docs/SouthamptonDAF.pdf (Accessed: 22 February, 2016).

Goff, M., Poschen, M., Procter, R., Halfpenny, P., Beard, L., Besson, J., Collins, S., Finch, J. et al. 2010. *The implications of disciplinary practices for emerging modes of data sharing : a case study of Biomedical researchers.* In: Proceedings of the 9th UK e-Science All Hands Meeting (AHM 2010), Cardiff, 13-16 Sep 2010. [PDF]. [Online]. Available: http://www2.warwick.ac.uk/fac/sci/dcs/people/rob_procter/publications/ahm2010_submission_99.pdf (Accessed: 12 February, 2016).

Halder, S. N. & Chandra, S. 2013. Essentials of Institutional Digital Repository in Academic Library: A Case Study. *Indian Journal of Information Sources & Services (IJISS)*, *3*(2):1-5.

Harris-Pierce, R.L. & Liu, Y.Q., 2012. Is data curation education at library and information science schools in North America adequate? *New Library World*, *113*(11/12): 598-613.

HEFCE. 2014. *TRAC Guidance*. [Online]. Available: http://www.hefce.ac.uk/funding/finsustain/trac/ (Accessed: 12 February, 2016).

Henderson, M. E. & Knott, T. L. 2015. Starting a research data management program based in a university library. *Medical Reference Services Quarterly*, *34*(1): 47-59.

Henty, M. 2008. *Dreaming of data: the library's role in supporting e-research and data management*. Presentation at Australian Library and Information Association Biennial Conference, 2 – 5 September 2008 Alice Springs Convention Centre, Alice Springs, NT Australia. [PDF]. [Online]. Available: http://apsr.anu.edu.au/presentations/henty_alia_08.pdf (Accessed: 26 February, 2016).

Higgins, S. 2008. The DCC curation lifecycle model. *International Journal of Digital Curation*, *3*(1): 134-140.

Higgins, S. 2012. The lifecycle model of data management. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 17-46.

Higman, R. & Pinfield, S. 2015. Research data management and openness: The role of data sharing in developing institutional policies and practices. *Program: electronic library and information systems*, *49*(4): 364-381.

Hiom, D., Fripp, D., Gray, S., Snow, K. & Steer, D. 2015. Research data management at the University of Bristol: charting a course from project to service. *Program: Electronic Library and Information Systems*, *49*(4):475-493.

Holdren, J.P. 2013. *Increasing access to the results of federally funded scientific research*. Memorandum for the heads of executive departments and agencies. Office of Science and Technology Policy, Executive Office of the President, Washington, DC. [Online]. Available:

www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (Accessed: 19 January, 2016).

Horton, L. & DCC. 2014. *Overview of UK institution RDM policies*. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies (Accessed: 11 February, 2016).

JISC. n.d.a. *Managing research data (JISCMRD)*. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx (Accessed: 23 January, 2016).

JISC. n.d.b. *Managing research data programme 2011-13*. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx (Accessed: 23 January, 2016).

Jones. S. 2012a. Research data policies: principles, requirements and trends. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 47-66.

Jones. S. 2012b. *Summary of UK research funders' expectations for the content of data management and sharing plans (Version 4.4)*. Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/sites/default/files/documents/resource/policy/FundersDataPlanReqs_v4%204.pdf (Accessed: 23 January, 2016).

Jones. S. 2013. Bringing it all together: a case study on the improvement of research data management at Monash University. DCC RDM Services case studies. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/developing-rdm-services (Accessed: 11 February, 2016).

Jones. S. 2014. CARDIO: collaborative assessment of research data infrastructure and objectives. [Online]. Available: http://www.slideshare.net/sjDCC/cardio-34772337 (Accessed: 23 February, 2016).

Jones, S., Ball, A. & Ekmekcioglu, Ç. 2008. The data audit framework: a first step in the data management challenge. *International Journal of Digital Curation*, *3*(2):112-120.

Jones, S., Guy, M. & Pickton, M. 2013. Research data management for librarians. DCC course booklet. Edinburgh: Digital Curation Centre. [Online]. Available: www.dcc.ac.uk/webfm_send/1245 (Accessed: 12 January, 2016).

Jones, S., Pryor, G. & Whyte, A. 2012. *Developing research data management capability: the view from a national support service*. In: Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES), Toronto, 1-5 Oct 2012. : 142-149. [PDF].[Online]. Available: https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf (Accessed: 15 February, 2016).

Jones, S., Pryor, G. & Whyte, A. 2013. *How to develop research data management services - a guide for HEIs. DCC how-to guides*. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf (Accessed: 20 October, 2015).

Jones, S., Duke, M., Rans, J. & Welgert, V. 2015. Meeting the requirements of the EPSRC research data policy. [Online]. Available: https://www.jisc.ac.uk/guides/meeting-the-requirements-of-the-EPSRC-research-data-policy (Accessed: 29 February, 2016).

Jones, S., Ross, S., Ruusalepp, R. & Dobreva, M. 2009. *Data Audit Framework Methodology (Version 1.8)*. Glasgow, HATII and Digital Curation Centre. [Online]. Available: http://www.data-audit.eu/DAF_Methodology.pdf (Accessed: 21 February, 2016).

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.

Kim, J. 2013. Data sharing and its implications for academic libraries. *New Library World*, *114*(11/12): 494-506.

Knight, G. 2012. A digital curate's egg: a risk management approach to enhancing data management practices. *Journal of Web Librarianship*, *6*(4):228-250.

Koltay, T. 2015. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 1-12 [Preprint].

Lavoie, B. 2012. Sustainable research data. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 67-82.

Leedy, P. and Ormrod, E. J. 2014. *Practical research: planning and design*. 10th ed. Edinburg: Pearson.

Lewis, J. A. 2014. *Research data management technical infrastructure: a review of options for development at the University of Sheffield*. [Online]. Available: http://files.figshare.com/1801108/RDMTI_v2.4.pdf (Accessed: 28 February, 2016).

Lewis, M. J. 2010. *Libraries and the management of research data.* [Online]. Available: http://eprints.whiterose.ac.uk/11171/ (Accessed: 1 October, 2015).

Lötter, L. 2014. *Reflections on the RDM position in South Africa*. Presentation at the LIASA Developing Research Data Management Services Workshop, Cape Town 27 March 2014. [PPT]. [Online]. Available: http://www.dcc.ac.uk/webfm_send/1631 (Accessed: 26 January, 2016).

Lynch, C. and Carleton, D.E. 2009. Lecture: impact of digital scholarship on research libraries. Journal of Library Administration, 49(3): 227-244.

Lyon, L. 2012. The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, *7*(1): 126-138.

Lyon, L., Rusbridge, C., Neilson, C. & Whyte, A. 2010. *DCC SCARP: disciplinary approaches to sharing, curation, reuse and preservation -- JISC final report.* [Online]. Available: http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf (Accessed: 26 January, 2016).

Meriel, P. 2010. *Use of the data audit framework within the Sudamih project*. [Online]. Available: http://sudamih.oucs.ox.ac.uk/docs/Use%20of%20the%20DAF.pdf (Accessed: 22 February, 2016).

NHMRC. 2016. *NHMRC statement on data sharing*. [Online]. Available: http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing (Accessed: 23 January, 2016).

NISO. 2004. *Understanding Metadata*. Bethesda, MD:NISO Press. [Online]. Available: http://www.niso.org/publications/press/UnderstandingMetadata.pdf (Accessed: 1 September, 2016).

OECD. 2007. *Principles and guidelines for access to research data from public funding, OECD*. [Online]. Available: http://www.oecd.org/science/sci-tech/38500813.pdf (Accessed: 26 January, 2016).

Ogier, A., Hall, M., Bailey, A. & Stovall, C. 2014. Data management inside the library: assessing electronic resources data using the data asset framework methodology. *Journal of Electronic Resources Librarianship*, 26(2): 101-113.

Ohemeng, F.L. and Ofosu-Adarkwa, K. 2015. One way traffic: The open data initiative project and the need for an effective demand side initiative in Ghana. *Government Information Quarterly*, *32*(4): 419-428.

Onyancha, O.B. 2016. *Open Research Data in Sub-Saharan Africa: a bibliometric study using the data citation index*. Paper presented at the Fourth CODESRIA Conference on Electronic Publishing, Darkar, Senegal, 30 March – 1 April 2016. [Online]. Available: http://www.codesria.org/spip.php?article2564&lang=en (Accessed: 19 August, 2016)

ORID. 2015. *University of Ghana: Office of Research, Innovation and Development: vision and mission*. [Online]. Available: http://www.orid.ug.edu.gh/about/vision (Accessed: 3 August, 2015).

ORID. 2016. *University of Ghana Research Report, 2014/2015*. [Online]. Available: http://www.orid.ug.edu.gh/annual-research-report (Accessed: 16 August, 2016).

Patton, M. Q. 2015. *Qualitative research and evaluation methods*. 4th ed. Los Angeles, CA: Sage.

Pennock, M. 2006. *Digital preservation: continued access to authentic digital assets.* JISC Digital preservation briefing paper. [Online]. Available: www.dcc.ac.uk/webfm_send/378 (Accessed: 12 January, 2016).

Pickard, A. J. 2008. *Research methods in information*. London: Facet.

Pienaar, H., 2011. *Survey of research data management practices at the University of Pretoria: undertaken by the Department of Library Services in order to improve research practices at the University*. Paper presented at the Unisa Library Open Scholarship Seminar, Bamboo Auditorium, Kgorong Building (KGO 3-006), Muckleneuk Campus, Unisa, Pretoria, 25 October 2011. [Online]. Available: http://repository.up.ac.za/handle/2263/17475 (Accessed: 19 January, 2016).

Poschen, M., Finch, J., Procter, R., Goff, M., McDerby, M., Collins, S., Besson, J., Beard, L. et al. 2012. Development of a pilot data management infrastructure for biomedical researchers at University of Manchester -- approach, findings, challenges and outlook of the MaDAM project. *International Journal of Digital Curation*, 7(2):110-122.

Poschen, M., Goff, M., Procter, R., Halfpenny, P., Beard, L., Besson, J., Collins, S., Finch, J. et al. 2010. *User-driven development of a pilot data management infrastructure for biomedical researchers*. In: UK e-Science all hands meeting. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.8161&rep=rep1&type=pdf (Accessed: 22 February, 2016).

Procter, R., Halfpenny, P., & Voss, A. 2012. Research data management: opportunities and challenges for HEIs. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 135-150.

Pryor, G. 2012. Why manage research data? In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 1-16.

Pryor, G. 2013. A maturing process of engagement: Raising data capabilities in UK higher education. *International Journal of Digital Curation*, *8*(2):181-193.

Pryor, G. 2014. A patchwork of change. In: Pryor G., Jones S. & Whyte A. (eds.), *Delivering research data management services: fundamentals of good practice.* London: Facet Publishing: 1-19.

Pryor, G. & Donnelly, M. 2009. Skilling up to do data: whose role, whose responsibility, whose career? *International Journal of Digital Curation*, *4*(2), pp.158-170.

Qin, J. 2013. Infrastructure, standards, and policies for research data management. In: *Sharing of scientific and technical resources in the era of big data*: The Proceedings of COINFO 2013, :214-219. Beijing: Science Press. [Online]. Available: http://0-surface.syr.edu.innopac.up.ac.za/istpub/164/ (Accessed: 19 January, 2016).

Raboin, R., Reznik-Zellen, R. C., & Salo, D. 2012. Forging new service paths: institutional approaches to providing research data management services. *Journal of eScience Librarianship*, 1(3): 134-147.

Rans, J. & Jones, S. 2013. *RDM strategy: moving from plans to action*. DCC RDM Services case studies. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/resources/developing-rdm-services (Accessed: 12 February, 2016).

Ray, J.M., 2014. Introduction to research data management. In: Ray, J. M. (ed.) *Research Data Management: Practical Strategies for Information Professionals.* Series: Charleston Insights in Library, Information, and Archival Sciences. West Lafayette, IN: Purdue University Press: 1-21.

RCUK. 2011. *RCUK common principles on data policy*. [Online]. Available: http://www.rcuk.ac.uk/research/datapolicy/ (Accessed: 19 January, 2016).

Rice, R. & Haywood, J. 2011. Research data management initiatives at University of Edinburgh. *International Journal of Digital Curation*, *6*(2):232-244.

RIN. 2008. *To share or not to share: publication and quality assurance of research data outputs -- report commissioned by the Research Information Network*. [Online]. Available: http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf (Accessed: 26 January, 2016).

Searle, S., Wolski, M., Simons, N. & Richardson, J. 2015. Librarians as partners in research data service development at Griffith University. *Program: electronic library and information systems*, *49*(4): 440-460.

Sproull, N.L. 2002. *Handbook of research methods: a guide for practitioners and students in the social sciences*. 2nd ed. Lanham, MD: Scarecrow.

Strasser, C., Cook, R., Michener, W. & Budden, A. 2012. *Primer on data management: what you always wanted to know*. [Online]. Available: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf (Accessed: 12 January, 2016).

Surkis, A. & Read, K. 2015. Research data management. *Journal of the Medical Library Association (JMLA)*, *103*(3): 154.

Takeda, K., Brown, M., Coles, S., Carr, L., Earl, G., Frey, J., Hancock, P., White, W. et al. 2010. *Data management for all: the institutional data management blueprint project*. Presentation at the 6th International Digital Curation Conference. [Online]. Available: http://eprints.soton.ac.uk/169533/1/6th_international_digital_curation_conference__idmb_final_paper_revised.pdf (Accessed: 29 July, 2015).

Tam, W., Fry, J. & Probets, S. 2014. *The disciplinary shaping of research data management practice*s. In: IConference 2014 proceedings. : 721-728. [Online]. Available: https://www.ideals.illinois.edu/handle/2142/47276 (Accessed: 22 January, 2016).

Treloar, A., Choudhury, G.S. & Michener, W. 2012. Contrasting national research data strategies: Australia and the USA. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 173-203.

UC3. 2011. *DMPTool.* [Online]. Available: https://dmptool.org/ (Accessed: 12 February, 2016).

UG. [2013]. *UG research policy guidelines: good practices: record keeping & data management*. [Online]. Available: www.orid.ug.edu.gh/sites/orid.ug.edu.gh/files/pictures/Guidelines%20on%20Good%20Practices.pdf (Accessed: 3 August, 2015).

UG. 2014. *University of Ghana Strategic Plan, 2014 – 2024*. [Online]. Available: http://www.ug.edu.gh/sites/default/files/documents/UG%20Strategic%20Plan.pdf (Assessed: 3 August, 2015).

UG. 2016. *University of Ghana*. [Online]. Available: http://www.ug.edu.gh/about/overview (Assessed: 16 August, 2016).

UK Data Archive. n.d. *Create and manage data: research data lifecycle*. University of Essex (2002– 2016). [Online]. Available: http://www.data-archive.ac.uk/create-manage/life-cycle (Accessed: 30 January, 2016).

UK Data Service. n.d. *UK Data Service*. [Online]. Available: https://www.ukdataservice.ac.uk/ (Accessed: 12 February, 2016).

University of Bristol. n.d. *Storing and using research data*. [Online]. Available: http://data.blogs.ilrt.org/usingdata/ (Accessed: 12 February, 2016).

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. & Horton, L. 2011. *Managing and sharing data: best practice for researchers. 3rd edition*. UK Data Archive: Essex. [Online]. Available: http://data-archive.ac.uk/media/2894/managingsharing.pdf (Accessed: 12 February, 2016).

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.

van Wyk, J. 2016. *Research data management (RDM): Subject guide*. Department of library services, UP. [Online]. Available: http://up-za.beta.libguides.com/c.php?g=356288&p=2404132 (Accessed: 12 February, 2016).

Wang, M. 2013. Academic Library, e-Science/e-Research, and data services in a broader context. In: *Proceedings of the ACRL 2013 Conference in Indianapolis, Indiana, April 10-13, 2013.* Association of College and Research Libraries. [Online]. Available: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/Wang_AcademicLibrary.pdf (Accessed: 12 January, 2016).

Welgert, V. 2015. *How can universities meet the expectations in the EPSRC research data policy?* [Online]. Available: https://www.jisc.ac.uk/blog/how-can-universities-meet-the-expectations-in-the-epsrc-research-data-policy-27-mar-2015 (Accessed: 20 September, 2015).

Weller, T. & Monroe-Gulick, A. 2014. Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, *32*(3): 467-482.

Whitmire, A.L., Boock, M. & Sutton, S.C. 2015. Variability in academic research data management practices: implications for data services development from a faculty survey. *Program: Electronic Library and Information Systems*, *49*(4): 382-407.

Whyte, A. 2012a. Emerging infrastructure and services for research data management and curation in the UK and Europe. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet Publishing: 205-234.

Whyte, A. 2012b. *Turning roadmaps to action*. [Online]. Available: http://www.dcc.ac.uk/news/turning-roadmaps-action (Accessed: 3 March, 2016).

Whyte, A. 2014. *Final results from the DCC RDM 2014 survey*. [Online]. Available: http://www.dcc.ac.uk/blog/rdm-2014-survey (Accessed: 26 January, 2016).

Whyte, A. & Allard, S. (Eds) 2014. *How to discover requirements for research data management services*. DCC How-to Guides. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/how-discover-requirements (Accessed: 21 February, 2016).

Whyte, A. & Tedds, J. 2011. *Making the case for research data management*. DCC briefing papers. Edinburgh: Digital Curation Centre. [Online]. Available: www.dcc.ac.uk/webfm_send/487 (Accessed: 12 January, 2016).

Whyte, A. & Wilson, A. 2010. *How to appraise & select research data for curation*. Edinburgh: Digital Curation Centre. [Online]. Available: http://www.dcc.ac.uk/resources/how-guides/appraise-select-data (Accessed: 16 February, 2016).

Whyte, A., Molloy, L., Beagrie, N., & Houghton, J. 2014. What to measure? Towards metrics for research data management. In: Ray, J. M. (ed.) *Research Data Management: Practical Strategies for Information Professionals.* Series: Charleston Insights in Library, Information, and Archival Sciences. West Lafayette, IN: Purdue University Press: 275-302.

Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., Litauer, R., Lots, K. et al. 2013. *Data management guide for public participation in scientific research. DataOne working group*: 1-15. [Online]. Available: https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf (Accessed: 12 January, 2016).

Williams, R., Pryor, G., Bruce, A., Macdonald, S., Marsden, W., Calvert, J., Dozier, M. & Neilson, C. 2009. *Patterns of information use and exchange: case studies of researchers in the life sciences.* [Online]. Available: http://rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf (Accessed: 12 January, 2016).

Wilson, J.A. & Jeffreys, P. 2013. Towards a unified university infrastructure: the data management roll-out at the University of Oxford. *International Journal of Digital Curation*, *8*(2): 235-246.

Wilson, J. A., Martinez-Uribe, L., Fraser, M. A. & Jeffreys, P. 2011. An institutional approach to developing research data management infrastructure. *International Journal of Digital Curation*, 6(2): 274-287.

Witt, M. 2008. Institutional repositories and research data curation in a distributed environment. *Library Trends*, *57*(2):191-201.

Wong, G. K. 2009. Exploring research data hosting at the HKUST institutional repository. *Serials Review*, *35*(3): 125-132.

Yin, R.K. 2009. *Case study research: Design and methods*. 4th ed. (Applied Social Research Methods Series vol. 5). Los Angeles, CA: Sage.

APPENDICES

Appendix A: Semi – Structured Interview guide

Dear Respondent,

I am a student at the University of Pretoria, South Africa currently studying for a Master of Information Technology degree. I am gathering data to enable me to complete my Masters Dissertation on the topic -- "**Developing an implementation plan for research data management (RDM) at the University of Ghana**".

The study will seek to gain an understanding of research data management requirements, activities and capabilities here at the University of Ghana (UG). I will, therefore, like to find answers to these questions:

1. Are there existing, well communicated, and well implemented policies, procedures and systems to ensure effective RDM at UG?
2. Are there adequate technologies (hardware, software and security infrastructure) to support effective RDM at UG?
3. Are there sufficient and sustainable resources (e.g. finance, human, skills and competence) to support effective RDM at UG?

RDM is "the active management and appraisal of data over the lifecycle of scholarly and scientific interest", and basically encompasses all activities that capture, document, organise, curate, preserve and provide continuous access to **research data** during and after the active research phase.

Please be assured that your identity in this interview will be handled with strict confidentiality. Thank you very much for agreeing to participate in this interview, I am most grateful for your time.

Sincerely,

Bright K. Avuglah

I. INTERVIEW GUIDE FOR A SENIOR OFFICIAL WITHIN UGLS (LIBRARY)

1. Can you please describe briefly what your role at the library is?

2. Please describe how the library currently supports the research process at UG. (*If response does not make reference to RDM, ask follow up question*)

2.1. What about providing support for the data generated by researchers? Is there any sort of training, support service or guidance (such as materials, workshops, audio or videos etc.) that is provided for researchers in terms of how to manage or share their data throughout the research lifecycle? (*If no support for RDM, ask follow up question*)

2.2. Does the library have any plan for implementing (setting up) services to support RDM anytime in the future?

3. Do you consider that there is the need for the University and the library for that matter to provide support for researchers of UG in terms of how they manage and share their research data?

4. In your opinion who should be responsible for RDM development and support at UG?

5. Can you please tell me if there are some staff here in the library who may be knowledgeable in aspects of RDM (e.g. providing training and guidance on data management, or knowledge in data curation, DMPs etc)? And if there are, are they known to researchers in the university community and are they consulted for assistance?

6. In your opinion, do you consider that the staff of the library possess the right knowledge and skills and are they well equipped to provide support for data curation and RDM at UG?

6.1. (**If yes to Q6**) Could you please elaborate on some of these knowledge, skills, and tools that currently exist in the library to support data curation and RDM?

6.2. (**If No to Q6**) What kind of training or tools would you require or suggest to be made available to be able to support RDM and data curation in the library?

7. In your opinion, would it be necessary for the University to identify and create a catalogue of important data assets within UG? And do you find the skills of the library staff relevant in this regards, particularly in the creation of metadata?

8. Does the institutional repository (IR) currently managed by the library permit deposit of datasets?

8.1. (**If yes to Q8**) are researchers actually depositing their dataset? Who is responsible for curating these data? Who determines which data to keep? And how long can the data be kept in the repository?

8.1.1. Are there policies guiding access to these data?

8.2. (**If no to Q8**) are there plans to accept data deposit in the future?

9. What challenges do you envisage in trying to support data storage through the IR?

10. Please is there any other information/opinion you will like to share with me?

II. INTERVIEW GUIDE FOR A SENIOR OFFICIAL WITHIN UGCS (I.C.T. DIRECTORATE)

1. Can you please describe briefly what your role at the ICT Directorate is?

2. Please describe how the University of Ghana Computing System (UGCS) currently supports the research process at UG.

3. How does the current IT infrastructure of the university support collaborative research? (*If response does not make reference to RDM, ask follow up question*)

3.1. What about providing support for the data generated by researchers? How adequate and secure is the current IT infrastructure (e.g. network bandwidth, storage) to support active data storage, regular and automatic backing up and syncing of data as well as facilitating data sharing?

4. Would you consider the current IT infrastructure robust and resilient enough to cope with an increase in demand for data storage space?

5. Does the UGCS have any data management platform or applications to assist researchers to manage their data effectively (including data storage and sharing)?

5.1. (**If no to Q5**) Is the University/UGCS considering developing a research data management platform for UG?

6. Is the University/UGCS also considering providing data storage service for researchers in the future?

7. Do you consider some of the UGCS staff to be knowledgeable about research data management and could support researchers in that regard?

8. Would you consider it necessary for the University to support researchers in the management of their research data (including data storage, sharing and archiving)? And do you think it can be done sustainably?

9. Can the university's IR support data deposit? And what will be required to make research data storage in the IR possible and sustainable?

10. Please is there any other information/opinion you will like to share with me?

III. INTERVIEW GUIDE FOR A SENIOR OFFICIAL AT ORID (RESEARCH OFFICE)

1. Can you please describe briefly what your role here at the Office of Research, Innovation and Development (ORID) is?

2. Can you please describe how ORID support researchers and research activities here at the University of Ghana?

3. What is the attitude of the University towards RDM? And how is the University enabling good data management practices at UG?

4. Is the University conversant with existing research data assets available within the University (their location and volume) and risk associated with poorly managing these research data? How are these data assets managed?

5. How do you determine which data are of long-term value and should be preserved or shared if possible?

6. How long do you expect researchers to keep their research data and how are you ensuring compliance?

7. How does the University address the governance of data access and re-use? To what extent do you promote open sharing of research data and not just the research findings?

8. How is the ORID ensuring that existing policies and guidelines on research and RDM are well communicated to the research community of UG and that they are adhered to?

9. The UG research policy states on page 6 "The University shall put in place a system for [among other things] managing research data/ material" and "the University will create a meta-database of research materials/ data repositories". Has the University succeeded in putting in place these systems and how has this been done?

9.1. (**If No**) what is hindering the setting up of these systems? What challenges have been encountered?

10. Would you consider it necessary for the University to provide support for researchers in terms of how they manage and share or in preserving their research data? And do you think it can be done sustainably?

11. Is there any sort of training, support service or guidance (such as materials, workshops, audio or videos etc.) that is provided for researchers in terms of how to manage or share their research data? (*If no support for RDM, ask follow up question*)

11.1. Is the University considering providing support for researchers in the management of their research data in the future? How does the university plan to do this and what areas will be considered?

12. Are researchers (both faculty and students) here required to write data management plans (whether as part of funding requirement or the research process) and how is the University supporting this requirement?

13. In your opinion, do you consider that some staff here at the ORID may be knowledgeable in aspects of RDM (e.g. providing training and guidance on data management)? And if there are, are they known to researchers in the university community and are they consulted for assistance?

14. Does the University have any resource allocations (including funding) to facilitate RDM support/ development in the near future?

15. Please is there any other information/opinion you will like to share with me?

IV. INTERVIEW GUIDE FOR RESEARCHERS

1. Can you please describe your role at the University of Ghana? How long have you been in this capacity, and what your discipline and area of research focus are?

2. What types of research data do you often generate in your research activities? How do you manage, store and preserve your research data during and after any research activity?

3. In terms of your research outputs, what contents are most important to you (e.g. research findings, article, data)?

4. Are you aware of existing institutional policies on research and specifically on research data management? Where and how do you assess these policies?

5. How long do you keep your research data before disposing it?

6. How do you share your research data? And how do you determine what data has long term value? What role does the university play in this regard?

7. Does the University of Ghana (through its IT services or library) offer secure storage and backup services for your research data? Have you used such a service before?

7.1. (**If NO to Q8**) do you consider such services or support as appropriate for you as a researcher?

8. How confident are you about existing IT infrastructure within the University to be able to support the growing volumes of institutional research data and storage needs?

9. Do you get external funding for your research? And have you been required at any time to write a data management plan – outlining how the data will be collected, managed and stored – as an addendum to the proposal for funding?

10. How does the University of Ghana (through the library, research office or any support unit) support you in writing data management plans, if you have had the need to write one?

11. Have you ever received any support from the university support services (such as the library, IT services or research office) in the form of training, workshops, audio or video materials specifically on issues of proper management of research data and good research practices?

11.1. (**If NO to Q12**) do you consider such services or support appropriate for you as a researcher?

12. Do you think it is important for the University to support researchers in the stewardship – continuous management, sharing, archiving and preserving – of their research data? And do you think it can be done sustainably?

13. Are you aware of the potential risks associated with poor research data management? And what kind of institutional systems and support are available to you here at the University of Ghana in curtailing such risks?

14. Please is there any other information/opinion you will like to share with me?

Appendix B: Ethical Clearance from University of Ghana

UNIVERSITY OF GHANA

ETHICS COMMITTEE FOR THE HUMANITIES (ECH)

P. O. Box LG 74, Legon, Accra, Ghana

My Ref. No.....

10th October 2016

Mr. Bright K. Avuglah
University of Pretoria
Pretoria

Dear Mr. Avuglah,

ECH 011/16-17: DEVELOPING AN IMPLEMENTATION PLAN FOR RESEARCH DATA MANAGEMENT AT THE UNIVERSITY OF GHANA

This is to advise you that the above reference study has been presented to the Ethics Committee for the Humanities for a full board review and the following actions taken subject to the conditions and explanation provided below:

Expiry Date:	26/03/17
On Agenda for:	Initial Submission
Date of Submission:	25/07/16
ECH Action:	Approved
Reporting:	Quarterly

UG/ISSER/ECH
VALID UNTIL
26 MAR 2017
APPROVED DOCUMENT

Please accept my congratulations.

Yours Sincerely,

Rev. Prof. J. O. Y. Mante
ECH Chair

Tel: +233-303933866 Email: ech@ug.edu.gh | ech@isser.edu.gh

Appendix C: Data Management Plan

Created with DCC Checklist for a Data Management Plan v4.0 (DCC, 2013d)

ADMIN DETAILS

Project Title: Developing an implementation plan for Research Data Management (RDM) at the University of Ghana, Legon

Researcher: Bright K. Avuglah

Affiliation: University of Pretoria

Supervisor: Emeritus Prof. Peter G. Underwood

Last edited: December 29, 2016

Description: The study is a qualitative case study which examines the state of the art of RDM development and institutional preparedness at UG through existing activities and capabilities. The data to be collected will help to understand institutional awareness, activities, and capabilities, and to answer important questions like whether there are existing and well-implemented policies and procedures on RDM, whether existing technical and technology capabilities are adequate and whether there are adequate resources (human, skill and financial) to support RDM at University of Ghana.

DATA COLLECTION

What data will you collect or create?

Qualitative data will be collected in audio format through interviews with seven respondents (5 service providers and 2 senior researchers). Owing to time constraints, the audio recordings will be partially transcribed so that only important parts that can be used for directed quotations in the data analysis will be transcribed and anonymised.

Existing data will also be gathered through institutional policies and guidelines particularly, relating to research and research data management. These documents are openly available on the website of the University of Ghana or the Office of Research, Innovation and Development (ORID). The volume of data is expected to be more or less 180 MB.

How will the data be collected or created?

The interviews will be recorded with the researcher's mobile phone (Infinix Note 2 X-600). The audio files will be created and stored in .amr format during the research. Microsoft Office Word 2007 will be used to partially transcribe the interviews and saved in .docx format. At the end of the study, the audio files will be converted into .mp3 format and the text file will be converted into .pdf (recordings and text files will be stored in both original

formats [.amr and .docx] and in the converted formats [.mp3 and .pdf] for long-term preservation).

All the research data (audio recordings and text files) will be stored in a folder labeled as Avuglah_MITDissertation_Recordings ("Avuglah" is the surname of the researcher and "MIT" stand for *Master of Information Technology* which is the programme for which this study was conducted as part of the requirement for the award of the degree). Each audio file will be named as: [respondent's code]_[location/unit of the respondent] e.g. (***R5_ORID***), and the text file will be named as Avuglah_MITDissertation_Transcript.

DOCUMENTATION AND METADATA

What documentation and metadata will accompany the data?

The entire interviews will be transcribed in one document under different headers which clarify specific metadata on them. Each header captures the following information in strict order to describe the particular interview or data: **Date: Respondent's Code: Respondent location/unit: Time (start - finish): Recording device used: a brief optional remark on the particularly data collection activity**. E.g. (Oct. 12, 2016: R1: UGCS: 10:00 am -10:30 am: Infinix Note 2 – X600: the participant was in a hurry to attend an official meeting, so we had to somewhat speed up the conversation). "UGCS" stands for University of Ghana Computing System.

All interviews will be conducted by the researcher and a semi-structured interview guide will be used to collect the data (each unit will have a unique set of questions to be asked)

ETHICS AND LEGAL COMPLIANCE

How will you manage any ethical issues?

The University of Ghana and the University of Pretoria have both given ethical clearance for this study to be conducted. Consent was also given by all participants to willingly participate in this study. As much as possible, the researcher will maintain the confidentiality of the participants in the analysis of the data by anonymising the data and using codes to refer to the participants.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Any intellectual property arising from the data collected in this study will be held by the University of Pretoria and the Student.

STORAGE AND BACKUP

How will the data be stored and backed up during the research?

Audio files will be stored in both .amr and .mp3 file formats and text files will be stored in both .docx and .pdf formats. Throughout the study and after completion the data will be stored on the student's computer hard drive and backup Google Drive (cloud).

DATA SHARING

How will you share the data?

Essentially, data will be shared through person to person exchanges once the person requesting for the data can proof and justify the need for the data.

Are any restrictions on data sharing required?

The data will be collected through interviews with participants and therefore making data available to other people will require some consent from the participants unless the data is required for inspection, audit or protection by the University of Pretoria, Faculty of Engineering, the Built Environment and Information Technology.

As it is possible for respondents to give out sensitive information during the interview which will be captured on the audio recording, it may be necessary in some cases to rather make available the transcripts where those sensitive portions will be edited before sharing. In other cases, the recipient must provide assurances that the identity of participants will be kept confidential.

RESPONSIBILITIES AND RESOURCES

Who will be responsible for data management?

For now, the researcher will be responsible for all the research data management activities, from data planning, collection, processing, analysis, storage and preservation, and sharing unless the University of Pretoria offers to take up any of such responsibilities.