AN EVALUATION OF A STRUCTURED TRAINING EVENT AIMED AT ENHANCING THE RESEARCH DATA MANAGEMENT (RDM) KNOWLEDGE AND SKILLS OF LIBRARY AND INFORMATION SCIENCE (LIS) PROFESSIONALS IN SOUTH AFRICAN HIGHER EDUCATION INSTITUTIONS (HEIs)

Mini-dissertation by

Refiloe Letticia Matlatse

29315370

Submitted in partial fulfilment of the requirements for the degree of

MASTER OF INFORMATION TECHNOLOGY (B)

in the

FACULTY OF ENGINEERING, THE BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

Supervisors: Dr. H. Pienaar & Dr. MJ van Deventer

January 2016

ABSTRACT

Research Data Management (RDM) has received a lot of attention recently. In South Africa, the importance of RDM has amplified since the release of the National Research Foundation's (NRF) open access statement. According to the statement, researchers who receive funding from the NRF must deposit their research output in an open access (OA) repository. In addition, the data supporting the research should be deposited in an accredited OA repository with a Digital Object Identifier (DOI) for future citations (NRF, 2015: online). The mandate, along with other drivers such as research data re-use, increased impact and validation of research findings has forced institutions to investigate the possibility of offering RDM services in their institutions (Ashley, 2012).

It is expected that libraries and Library and Information Science (LIS) professionals will initiate and support RDM in their institutions. LIS professionals will need to upgrade or obtain new skills and knowledge to fulfil their new roles and responsibilities. Various training opportunities are available to interested professionals to improve their knowledge and skills related to RDM. These can be as simple as a workshop or as complex as a university degree.

The objective of this research was to identify and evaluate a RDM training intervention to determine whether the training intervention could enhance the knowledge and skills of LIS professionals in South African (SA) Higher Education Institutions (HEIs). An embedded research design was used to investigate whether an RDM workshop, hosted by the Network for Data and Information Curation Communities (NeDICC), could enhance the LIS professional's (participants) perception of their RDM understanding, knowledge and skills. The research found that the RDM workshop was highly successful in enhancing the participant's perception of their RDM understanding and knowledge. The RDM workshop was less successful in enhancing the participant's perception of their RDM skills.

It was recommended that LIS professionals (1) take advantage of the online RDM training material available to enhance their understanding and knowledge of RDM; (2) attend face-to-face training interventions to enhance or develop their RDM skills and (3) enrol in university level educational programmes to gain a qualification in RDM if they qualify. It was also recommended that institutions that provide RDM training should focus on specific aspects of RDM instead of offering a general overview. This research can be used to inspire larger studies or studies that compare two or more RDM training interventions.

DECLARATION

I declare that the mini-dissertation, *An Evaluation of a structured training event aimed at enhancing the Research Data Management Knowledge and Skills of Library and Information Science Professionals in South African Higher Education Institutions,* is my own original work. When secondary material is used, this has been acknowledged and referenced in accordance with the University of Pretoria's requirements.

---

25 January 2016

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for giving me the opportunity and the strength to complete this research project. My sincerest thanks to my supervisors, Dr Heila Pienaar and Dr Martie van Deventer, for their patience and support throughout my studies. I am grateful to have worked with them.

I would like to thank the Carnegie Corporation of New York and the University of Pretoria for the financial assistance they provided. A special thank you to my employer for giving me the time I needed and to my colleagues for their support during my studies.

I want to thank my mother, Refilwe Letwaba and my sister, Basetsana Mothibe, for their continued support and for believing in me even when I had doubts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

LIST OF ABBREVIATIONS AND ACRONYMS

DCC - Digital Curation Centre

DMP - Data Management Plan

DOI - Digital Object Identifier

HEI - Higher Education Institution

ICT - Information and Communication Technology

IT - Information Technology

JISC - Joint Information Systems Committee

LIS - Library and Information Science

NeDICC - Network of Data and Information Curation Communities

NRF - National Research Foundation

OA - Open Access

RDA - Research Data Alliance

RDM - Research data management

SA – South Africa

UK- United Kingdom

US – United States

CHAPTER 1: INTRODUCTION

1.1. Background to RDM

The subject of research data management (RDM) is very well defined in the literature. The volume at which research data is produced due to advancements in technology has created the need for the management of research data (Cox & Pinfield, 2014: 299; Ray, 2014: 1 and Davidson *et al.,* 2014: 215). Research funding bodies have realised that research conducted using public funds belongs to the public. For this reason, it should be made accessible to the public, which would promote the need to manage research data (Chartered Institute of Library and Information Science (CILIP), 2014: 4; Tenopir *et al.,* 2014: 84; Kruse & Thestrup, 2014: 316 and van Wyk, 2014: online). In Higher Education Institutions (HEIs) it is anticipated that that libraries will take on the responsibility of initiating and supporting the management of research data. Lyon (2012a: 127) notes that although libraries have long supported scholarly communication and research by providing access to and preserving research output, they will have to re-examine their service portfolio to ensure they can support the data-driven research environment.

Lyon (2012a: 129-130) further provides examples of services that a library can offer. These include an RDM needs assessment and planning, and offering services such as RDM informatics, licencing, citation, training, appraisal, storage, access and measuring impact. A library should not take up this task in isolation. They should rather collaborate with other campus entities including the research office, the IT department, and the academic departments (Cox & Pinfield, 2014: 301; Kruse & Thestrup, 2014: 312 and Chad & Enright, 2014: 152).

Many challenges inhibit the development of RDM, one of them being the shortage of skilled professionals (Cox & Pinfield, 2014: 308; CILIP, 2014: 9 and Tenopir *et al*., 2014: 85 and van Deventer & Pienaar, 2015: 34). The LIS professionals who will be working on RDM in their institutions will be need to have certain knowledge, understanding and skills related to RDM (Auckland, 2012: 34 and Daland, 2015: 4).

Because RDM is still in its infancy, there is an opportunity for LIS professionals to enrol in educational programmes and attend training events to enhance their knowledge, understanding and skills in this regard (Auckland, 2012: 60 and Charbonneau, 2013: 366).

The study sought to investigate whether a structured training intervention can enhance the RDM knowledge, understanding and skills of LIS professionals. The problem statement and research questions are presented in section 1.2 that follows.

1.2. Problem statement and research questions

This section describes the problem statement of the research as well as the main research question and related sub-questions.

1.2.1. Problem statement

Over the past few years, research funding bodies in countries such as the United Kingdom (UK), the United States of America (USA), Canada and Australia have issued statements, calling for the management and sharing of publicly funded research data (Richardson *et al.,* 2012: 260; Davidson *et al.,* 2014: 216 and Tenopir *et al.,* 2014: 84). To take it a step further, the research funding bodies also require that researchers applying for funding submit a data management plan, outlining how they will manage the data they generate during the research project (Davidson *et al.,* 2014: 216 and Ray, 2014: 9).

In 2015, the National Research Foundation (NRF) in South Africa (SA) released an Open Access (OA) statement. In this statement, the foundation mandates that research output from research funded by the NRF must be deposited in an accredited OA Institutional Repository (IR). Furthermore, the data generated during the research project is required to be deposited in an OA repository with a Digital Object Identifier (DOI) (NRF, 2015: online).

SA HEIs will need to offer RDM services to ensure that their researchers comply with this mandate. Some of the challenges that they may face in offering these services will be developing infrastructure and upskilling the LIS professionals to prepare them for their new role. This research investigated if a specific RDM training event will enhance the LIS participants' RDM understanding, knowledge and skills.

1.2.2. Central research question

The central research question that the study sought to answer in the following:

To what extent would a structured RDM training intervention enhance the RDM understanding, knowledge and skills of LIS professionals at HEIs?

Research sub-questions

The related sub-questions that the study seeks to answer are the following:

i. What is seen internationally as RDM?
ii. Should libraries and LIS professionals play a role in their HEI's RDM efforts? What could these roles entail?
iii. What is the best way to obtain or improve their knowledge and skills of RDM?
iv. To what extent would a structured RDM training intervention enhance the RDM understanding, knowledge and skills of LIS professionals at SA HEIs?

In an effort to answer the research question and related sub-questions, a research methodology was considered. The research methodology is briefly discussed in the next section.

1.3. Research methodology

In empirical research, qualitative and quantitative research is used. In qualitative research, the researcher explores the attitudes, behaviours and experiences of the research subjects, whereas in quantitative research the researcher generates statistics through large-scale surveys (Dawson, 2009: 14-15). It is possible to combine the two methods in a single research project called "mixed methods research". Creswell (2013: 217) defines "mixed methods research" as the collection and analysis of qualitative and quantitative data. This research was conducted under the qualitative research paradigm.

To test the research hypothesis and/or answer the research question, a research design is a strategy used to guide the research (Hofstee, 2006: 113). The different research designs available to an empirical researcher include phenomenology, grounded theory, case study, experimental and quasi-experimental research designs (Kumar, 2011: 106 and Leedy & Ormrod, 2010: 137). The researcher found that an embedded research design would be most suitable research for this study. The embedded design combined a quasi-experimental research design (quantitative) and case study research design (qualitative).

Experimental research designs allow researchers to test a hypothesised relationship between a dependent and independent variable by manipulating the independent variable (Kirk, 2013: 6). In a case study research design, it is assumed that the case being studied is atypical and can provide insights into the events and situations prevalent in the group from where the case was drawn (Kumar, 2011: 126).

The research was designed around a RDM workshop hosted by the Network of Data and Information Curation Communities (NeDICC). The workshop was used as the training intervention. Attendees of the workshop became the research participants. Data was collected using two, similar online questionnaires created using Google Forms. The first questionnaire was sent before the workshop to assess the research participants' perceptions of their RDM understanding, knowledge and skills. The second questionnaire was sent after the workshop to determine if the research participants' perceptions of their RDM understanding, knowledge and skills had changed as well as to obtain feedback on the training intervention itself.

The results of both questionnaires were analysed and compared. This, together with the literature study, enabled the researcher to answer the research question and sub-questions. The main reason for the questionnaire before and after the workshop (the intervention) was to determine if the workshop made any difference to a participant's perception of his/her knowledge of RDM and the skills necessary to support RDM.

The research paradigm, design and data collection methods are discussed in more detail in Chapter 3. The justification for the research is discussed in the section that follows.

1.4. Justification for the research

International trends regarding RDM in HEIs have allowed the libraries and LIS professionals to predict that South Africa (SA) will follow in these trends. Managing research data is no longer optional because research-funding bodies are insisting that recipients of research funding develop a plan to manage their data and to make the data publicly accessible.

This study investigated whether a structured RDM training event could enhance the knowledge and skills of LIS professionals deployed at HEIs. Recommendations on how the LIS professionals can further improve their RDM knowledge and skills were developed and reported in Chapter 5. Although the results of the research cannot be generalised, HEIs and their libraries could use the recommendations provided to identify or develop RDM training for their employees.

1.5. Limitations and delineations of the research

This research project had the following limitations:

Geographic location

The study was limited to individuals attending the NeDICC RDM workshop. Although the attendees were from various institutions across the country, the majority of the attendees were representative of the Gauteng region.

Soft skills

The study did not consider the soft skills required for enabling RDM in HEIs. The focus was primarily on the technical and informational skills and knowledge of the LIS workforce.

Timing

The workshop was scheduled one week before the International Federation of Library Associations and Institutions (IFLA) annual conference took place. The timing of the workshop may have also affected the response rate to the second questionnaire as most of the participants were on conference leave.

In the case of the researcher's institution, some individuals who were going to the IFLA conference were asked not to attend the training event to give other individuals a chance of attending a training event.

Generalisation

The results of the study are only applicable to the participants of the workshop; therefore they cannot be generalised.

1.6. Value of the research

Numerous studies have been conducted on RDM internationally – Lyon (2007), Cox *et al.* (2012), Ishida (2014), Davidson *et al.* (2014). In SA, the literature on the subject of RDM is sparse. This study investigates if a training event will enhance the RDM knowledge and skills of the LIS professionals at HEIs.

The data collected from the research will, firstly, identify the (perceived) knowledge and skills gaps of the participants and determine if the RDM workshop has resulted in a change in those perceptions. The results of the research as well as the recommendations provided will (1) contribute to the body of literature about RDM in South Africa and (2) inspire further studies on the subject. This study can be seen as exploratory, as this is the first study of its kind in South Africa. Further studies are needed to extend the findings to the academic LIS profession in SA.

1.7. Clarification of key terms

This section defines terms used often throughout the study.

E-Research

E-Research is defined as computationally intensive, large-scale, networked and collaborative forms of research across all disciplines (Association of Research Libraries, 2014: online). For this research, e-Research is considered to be research activities, including research collaboration and research data management, performed using ICT-tools across multiple disciplines. In the literature study, the term E-Research is used interchangeably with the term E-Science.

Higher Education Institutions (HEIs)

In SA, "Higher Education Institutions" (HEIs) refers to universities and universities of technology (Universities South Africa, 2015: online). For this research, the abbreviation HEIs includes universities, universities of technology as well as research organisations and research councils.

Research Data

Research data refers to analogue or digital output from any systematic investigation including observational, experimental, simulation and derived or compiled data (CILIP, 2014: 4).

Research Data Management (RDM)

RDM refers to the manner in which research data is organised, structured, stored and maintained (University of Oxford, 2014: online). For this research RDM is considered to be the process of actively managing the data collected and used throughout the research lifecycle, to validate research findings or outcomes.

1.8. Division of chapters

The research report is divided into five chapters. These are as follows:

Chapter 1: Introduction

Chapter 1 provides the background to the study, outlining the value, justification and limitations of the entire research project/study.

Chapter 2: Literature Review

Chapter 2 reports on a review of the literature published on the subject. Central themes relevant to the study include: RDM, e-Research, e-skills and the role of the LIS in supporting research data management.

Chapter 3: Research Methodology

Chapter 3 describes the research methodology. The research paradigm, research design, sampling methods and data collection will be discussed in detail.

Chapter 4: Data Analysis

In chapter 4, the data collected is analysed and the results presented.

Chapter 5: Recommendations and Conclusion.

Chapter 5 concludes the research. The research questions are evaluated within the context of the study.

List of references/bibliography

All the sources cited throughout the document were listed in this section.

Appendices

The appendices include materials that do not fall under any of above mentioned chapters, as well as the tools used for data collection. These include the questionnaires and the programme from the training intervention.

1.9. Conclusion

This chapter provided the background to the study and introduced the research problem that was investigated. The justification, value and limitations of the study were discussed.
Relevant definitions and a list of abbreviations were also presented in this chapter. The next chapter reviews the literature published on the subject of RDM.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

As was indicated in section 1.2.1, the recently released statement by the NRF has intensified the importance of RDM in SA. The NRF mandates that research output from research funded by the NRF must be deposited in an accredited OA Institutional Repository (IR). Furthermore, the data generated during the research project is required to be deposited in an OA repository with a Digital Object Identifier (DOI) (NRF, 2015: online).

Research funding bodies in developed countries such as the United States of America (USA), the United Kingdom (UK) and Australia released similar statements several years ago (Davidson *et al.,* 2014: 216; Kahn *et al.,* 2014: 298 and Tiwari & Chand, 2014: 241). Governments in the above mentioned countries are/have been tasked with the challenge of developing an e-Research infrastructure to help manage research data and make it publicly accessible (Richardson *et al.*, 2012: 261 and Kahn *et al.*, 2014: 296). HEIs in the above mentioned counties are therefore expected to align their strategies with their national/government strategies (Richardson *et al.,* 2012: 261). If SA universities are to similarly engage in RDM, it is then important to consider the knowledge and skills required to enable RDM at HEI libraries.

This chapter explores the literature available on the subject of RDM, focusing on the roles and responsibilities, skills required to enable and support RDM in HEIs. Furthermore, the different training opportunities available to enhance RDM skills and knowledge will be identified and their content will be explored. To start, an overview of RDM is provided.

2.2. Research Data Management: An Overview

Research has entered the collaborative, computational and data-intensive fourth paradigm brought on by the integration of Information and Communication Technologies (ICT) in the research process/lifecycle (Tenopir *et al.,* 2014: 84). Researchers' use of new technologies throughout the research process result in larger amounts of data being generated at a rapid rate, a phenomenon referred to as the "data deluge" (Pryor, 2012: 2 and CILIP, 2014: 1). This data is produced in electronic format (Pryor, 2012: 2 and Thompson *et al.,* 2014: 843). Data that is not born digital is usually digitised to increase access to it (Thompson, 2014: 843). Typical data formats include prints, photographs, video and audio recordings, biological samples, databases and digital files (Higgins, 2012: 19 and Thompson, 2014: 847).

Research data is heterogeneous, which means that different disciplines produce different types of data, in various sizes (Kennan *et al.,* 2012: 60).

RDM is defined as "the process of storing and curating data generated from research" (Kennan *et al.,* 2012: 62). Data is organised from its creation at the beginning of the research lifecycle, through to the dissemination and archiving of the research output (Whyte & Tedds, 2011: 1). RDM involves a list of activities and processes undertaken to organise research data throughout the research cycle – from the start of the research project up until dissemination of the research output (Whyte & Tedds, 2011: 1). Corti *et al.* (2014: 17) add that the life span of research data often supersedes that of the research project.

A lifecycle model is often used to describe the entire RDM process (Carlson, 2014: 65). To gain a better understanding of RDM it is perhaps necessary to first gain an understanding of the broader concept 'research'. In the section that follows, the research lifecycle is discussed briefly.

2.2.1. The Research Lifecycle

The research lifecycle is non-linear, cyclical process. Theoretically, the research lifecycle consists of at least six (6) data-centric stages as illustrated below. In practice, however, it is not as simple (Pryor, 2012: 6) and there may be many variations. Figure 2.1 is used depict a basic research lifecycle.

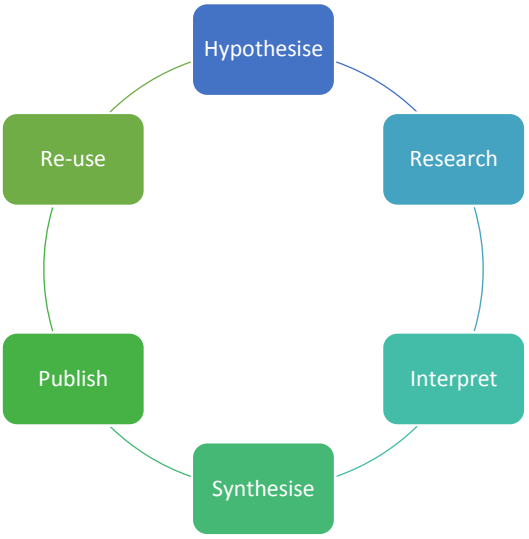


Figure 2.1: Research Lifecycle

(Source: Pryor, 2012: 6)

Hypothesise: at this stage, the researcher(s) plan for the research project (Pryor, 2012: 6 and Pickard, 2013: 79). The researchers may collect published data to review for their research project (Pryor, 2012: 6).

Research: researcher(s) generate new data through investigation or experimentation or other research methods. Researchers may choose to collect more (published) data for comparison or re-analysis (Pryor, 2012: 6; Higgins, 2012: 18 and Pickard, 2013: 79).

Interpret: data gathered through the investigation or experiment is processed and analysed (Higgins, 2012:18).

Synthesis: statistical techniques are used to make sense of the data and ultimately create new data/datasets (Higgins 2012: 18).

Publish: research data generated during the research stage is published and shared with others through scholarly books; articles; reports and other types of research output (Pickard, 2013: 79).

Re-use: published research may be accessed and used by other researchers. The researchers may use the published data to plan a new research project (hypothesise) or to compare with their own findings (Pryor, 2012: 6).

Akers & Doty (2012: 16) stress the importance of understanding the research lifecycle in its entirety before attempting to understand the research data lifecycle, which is discussed in the next section.

2.2.2. The Research Data Lifecycle

In the past, research data was considered a by-product of research output, but now research data is viewed as a commodity, a valuable product of research (Davenport & Patil, 2012: 72; Pryor, 2012: 6 and Higgins, 2012: 18). As with the research lifecycle, there isn't a single research data lifecycle model to suit all research projects. The model below is one example of a research data lifecycle. The model was created by the UK Data Archive and was seen as the most appropriate to illustrate the basic research data management concepts.

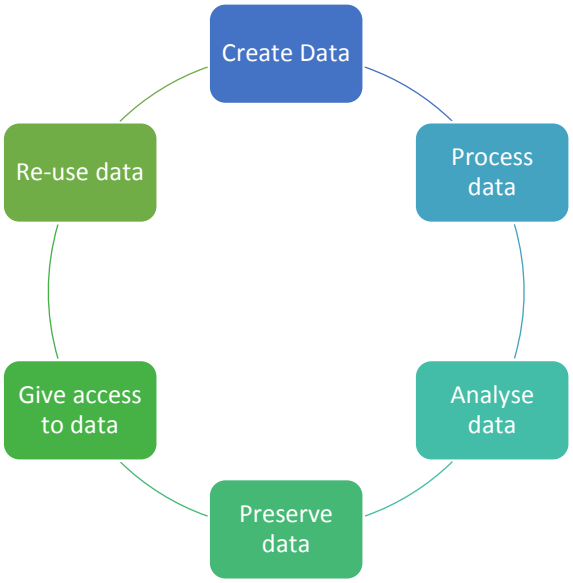


Figure 2.2: Research data lifecycle

(Source: UK Data Archive, 2014: online and van Wyk & van der Walt (2014: 5)

An expanded version of this model was created by Tiwari and Chand (2014: 242) and will be described and used in the section 2.2.3.

Research data must be managed as soon as it is created to protect it against the rapid advancement of technology, as well as other risks that may cause the data to become inaccessible or irretrievable (Higgins, 2012: 18).

According to Carlson (2014: 71), there are three (3) different types of research data lifecycle models: individual-based-, organisational-based- and community-based- research data lifecycle. The individual-based model is designed for the individual research project, the organisational-based model is designed specifically for the organisation/institution and the community-based model is designed for a specific community or discipline (Carlson, 2014: 71-75). It is up to the research team, the institution or the community to use a model that is most suitable to them.

The RDM lifecycle is discussed in section 2.2.3 below.

2.2.3. The Research Data Management Lifecycle

Authors such as Higgins (2012) and Pryor (2014) and Carlson (2014) refer to the Digital Curation Centre's (DCC) lifecycle model when describing RDM. Higgins (2012: 17) states that in information/archives/records management, a lifecycle model is used as it helps

counteract potential challenges. The lifecycle model can be developed to manage research data to help institutions cope with the changing technological landscape (Higgins, 2012: 17 and Clements, 2013: 301).

Tiwari and Chand (2014: 242) provide the eight stage RDM lifecycle model illustrated in figure 2.3 below.

Figure 2.3: Research Data Management Lifecycle

(Source: Tiwari & Chand, 2014: 243)

Tiwari and Chand (2014: 242) do not provide an in-depth description of the phases; instead, they identify potential roles where LIS professionals can take part in the lifecycle. The roles and responsibilities will be discussed later in the chapter. In brief, each of the stages could be described as follows:

Planning: planning for research data can begin as early as the grant writing stage (Tiwari and Chand, 2014: 243). Planning for research data includes designing the research; planning for data management, consent for sharing, collecting existing data sources (Corti *et al.*, 2014: 17).

Collecting: at this stage, data is collected through methods like experiments, simulation and observations. Existing third-party data may also be collected (Corti *et al.*, 2014: 17). Metadata is captured at this stage of the lifecycle (Corti *et al.*, 2014: 17).

Processing: data is entered, digitised, transcribed or translated if necessary (Corti *et al.,* 2014: 17). The data is also checked, validated, cleaned, possibly anonymised (Corti *et al.,* 2014: 17).

Analysing: at this stage of the RDM lifecycle the data is analysed and interpreted to produce research outputs (Corti *et al.*, 2014: 18).

Managing: file formats and mediums of storage are considered at this stage (Corti *et al.*, 2014: 18). It may also be necessary to document and organise the research data

Preserving: data is migrated to the best-format, and in the most suitable medium available (UK Data Archive, 2014: online). Metadata is created for the research data for the data to be archived (Corti *et al.*, 2014: 18 and UK Data Archive, 2014: online).

Sharing: copyright of the data should be established. The data can be shared with the research community however, the owner of the data can control access to the data (Corti *et al.,* 2014: 18 and UK Data Archive, 2014: online).

Re-using: this stage of the research entails secondary analysis of data, undertaking follow-up research, conducting research reviews and scrutinising research findings (Corti *et al.,* 2014: 18).

An example of successful alignment of RDM as well as data curation activities to research lifecycle activity is illustrated in the DCC lifecycle model (Proctor *at al.,* 2012: 139). Although the DCC lifecycle model provides a sequential illustration of the stages involved in the curation and preservation of data while also helping identify the roles and responsibilities of the stakeholders at each stage (Pryor, 2014: 41). It was not seen as essential for this research to discuss the DCC lifecycle model as the range of skills required could then become too broad.

2.2.4. The relationship between the research process and research data lifecycle

Figure 2.5, developed by Pickard (2013: 80), illustrates the relationship between the RDM cycle and the research process. The data management activities Data Management Planning (DMP), documentation, storage, appraisal, preservation and ingest) occur at the same time as activities of the research process (plan, collect, process and analyse, publish and new project). In this model, the researcher is responsible for activities in the lighter grey areas

(Pickard, 2013: 80). Other stakeholders, including libraries/LIS professionals are responsible for the areas in dark grey (Pickard, 2013: 80).

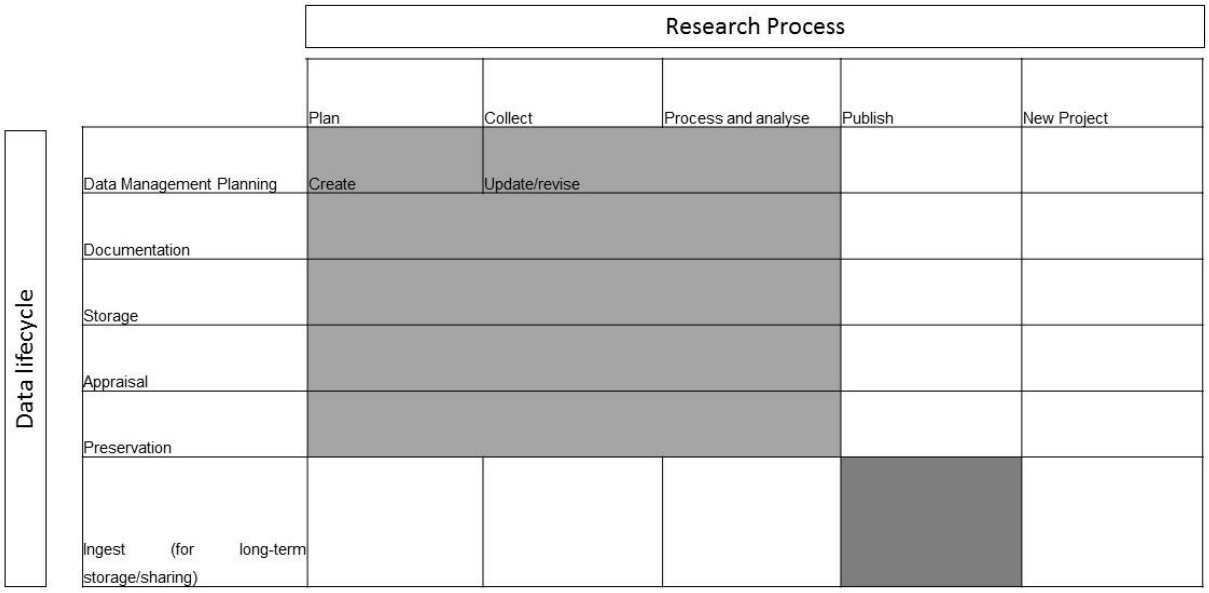


Figure 2.4: Relationship between the research process and the data lifecycle

(Source: Pickard, 2013: 80)

It is important to point out that different disciplines conduct research using different lifecycle models and processes. Regardless of the model used, research data should be managed throughout the research process and the output and data should be published to inspire new research projects.

The three lifecycle models depicted in Figures 2.2.1; 2.2.2. and 2.2.3 are all interrelated. For example the RDM lifecycle is based on a research data lifecycle which is based on a research lifecycle. Of course, these lifecycle models may vary depending on the project, the research project, the discipline or the institution or organisation within which the research is conducted. Using a lifecycle model to manage research data may become a problem when conducting collaborative or interdisciplinary research.

The drivers for successful RDM are discussed in the section that follows.

2.3. Drivers and benefits of successful research data management

Actively managing research data ensures that high-quality research is produced and that research data can be re-used for other research projects (University of Edinburgh Knowledge Strategy Committee, 2008 cited in Rice & Heywood, 2009: 6). HEIs, in compliance with research funding bodies, have developed or are in the process of developing data management policies and mandates (Corti *et al.,* 2014: 9). The following are seen as the key drivers for successfully managing research data.

2.3.1. Research data re-use

Data re-use is considered the primary reason to manage research data (Ashley, 2012: 155 and Cox & Pinfield, 2014: 302). This is not to say that all researchers should retain all data collected or generated during the research project because maintenance and storage costs may be incurred. Ashley (2012: 155) recommends that the cost of retaining the research data should be compared to the cost of generating new data to make informed decisions.
Further motivation for managing research data is that research is enhanced and validated by the availability of the data behind it (Ashely, 2012: 156, Davidson, 2014: 92 and Tiwari & Chand, 2014: 241). Research data shared should have accompanying, explanatory documentation about its production, processing, analysis and workflows.

2.3.2. Funder mandates and policies

According to Ashley (2012: 156), research funding agencies have realised the importance of managing and sharing publically funded data. There appears to be consensus that the data generated during the research projects belongs to the public that provided the funding. Many research funding agencies have issued mandates compelling researchers to make research data publicly available (Kahn *et al.,* 2014: 297).

The Research Councils UK (RCUK), for example, require that all research funded by the Council should be published in open access journals and a statement should be included stating how related research data can be accessed (Corti *et al.,* 2014: 5). Some of the funding agencies place the responsibility of managing and sharing research data with the researchers or research teams while others place the responsibility with the researcher's institution (Davidson *et al.,* 2014: 216).

Most of the UK RDM mandates issued require a data management plan (DMP) with the grant application to demonstrate how the data will be managed. A DMP helps design, practice and follow up on how research data should be collected, organised, used and maintained to achieve high quality and long term sustainability (Corti *et al.,* 2014: 24).

As was mentioned in section 2.1., the NRF is actually requesting seven actions. It urges researchers to (1) deposit supporting research data in an accredited open access repository, with the DOI to (2) enable access and citation (NRF, 2015: online and van Wyk, 2015: 2-3). The NRF further urges its stakeholder community to (3) formulate policies on open access publications and research data generated from funded research; (4) establish open access repositories; and (5) support public access to the repositories (NRF, 2015: online). In the past the NRF had encouraged the RDM community to (6) ensure data quality and to (7) facilitate data re-use (Kahn, *et al.,* 2014: 297). These actions already provide an indication of the skills that may need to be developed if this key driver is acknowledged as such.

2.3.3. Improving research impact

According to Brown *et al.* (2015: 18), some publishers already require authors to provide access to datasets with their research papers. By publishing high-quality research data, researchers are more likely to use the data. The data can then be cited or referenced by these other researchers, therefore improving the research impact of the original research (Ashley, 2012: 156 and Brown *et al.,* 2015: 18). Most libraries can provide assistance in tracking research output and impact using bibliometric analysis services such as the h-index (Richardson *et al.,* 2012: 266) – which again points to associated skills to be developed.

2.3.4. Authenticating research data

Researchers who manage their research data from the beginning of a research project will be able to make that data available to others who may question or doubt the research findings. Ensuring that the data can be located, retrieved and accessed is important when attempting to verify the research findings (Lyon, 2012b: 21 and Corti *et al.,* 2014: 1).

2.3.5. Other drivers for managing research data

Lyon (2012b: 19-24) identified additional reasons for managing research data. These are reducing the risk of losing or accidentally destroying the data; protecting the reputation of the researcher's institution because the underlying data will be available and accessible; and the

quality of research can be improved because research can be replicated and reproduced if data is available. The scale of research conducted could be increased because of the amount of data available and, finally, researchers will be able to build partnerships through interdisciplinary collaboration.

The drivers mentioned above provide good reason for RDM. However, RDM will only be effective if all the associated roles and responsibilities are clearly defined, mandated, and implemented.

2.4. Higher Education Institutions (HEIs): Roles and responsibilities

Many of the research-funding agencies have placed the RDM responsibility with the researcher's HEI (Davidson *et al.,* 2014: 216). Many of these HEIs have not yet established RDM services. For this reason the RDM roles, responsibilities and practices have not been established. Within the HEI, there are several key role players that can take on this responsibility. Pryor (2014: 45) identifies three groups within an HEI that may lead the RDM initiative, these are: HEI management; support and administrative services and the researchers or producers of the data. Their roles and responsibilities are described below.

2.4.1. The executive committee or university management

HEIs are in the process of developing strategies in response to the need for management of research data (CILIP, 2014: 7). In most HEIs, the Deputy Vice Chancellor (DVC) of Research is tasked with research planning and management (Richardson *et al.,* 2012: 264). RDM will then fall under the portfolio of the DVC of Research. Having full support from management cannot be overstated when initiating RDM, as considerable investment will be required (Pryor, 2014: 46). The investment will be used to develop infrastructure, recruit new staff or re-skill existing staff to offer RDM support (Pryor, 2014: 46).

2.4.2. Academic departments and researchers

Coates (2012: 53) suggests that researchers do not have the skills or desire to manage and share their research data. Kruse & Thestrup (2014: 317) mention that researchers may be discouraged to manage and share the research data they produce because of increased competition and lack of incentives. However, an earlier study by Akers & Doty (2012: 19) revealed that (early-career) researchers are willing to share their research data, but they lack the time to manage and share data.

The researchers will be the users of the RDM services provided by the HEI. Therefore, researcher buy-in is also mandatory to ensure researchers make use of the services and comply with their institutional policies and/or funder's mandates (Pryor, 2014: 50).

The research by Akers & Doty (2012: 18) revealed that researchers would use the following RDM-related services from their institutions:

- Workshops focusing on data management practices
- Assistance in preparing data management plans
- Consultation on data confidentiality and legal issues
- Personalised consultation on research data management for specific research projects/groups
- Institutional repository for research data
- Assistance with data documentation/metadata creation
- RDM workshops for trainees
- Digitisation of paper-based/physical research objects
- Assistance identifying discipline specific data repositories
- Data citation methods

(Akers & Doty, 2012: 18)
The services listed above could of course all be linked to skills development initiatives.

2.4.3. Support departments

The support and administrative service departments that will most likely be involved in providing RDM services will be the library, the IT/ICT department, research support and the repository or records management division (Proctor et al., 2012: 140 and Pryor, 2014: 48). In some HEIs, such as the University of Pretoria (UP), the repository division functions as part of the library. Student administration is responsible for records management.

The library and the IT/ICT department will have a crucial role to play in implementing RDM in their institutions. These two support service departments will have to work closely together when developing the RDM service infrastructure (Proctor et al., 2012: 140; Clements, 2013: 299 and Pryor, 2014: 48).

These roles and their associated responsibilities are unpacked in Table 2.1.

Table 2.1: Roles and associated responsibilities within HEIs

HEI Management	**Researchers and Academic Departments**	**Support Departments**
Establish a working group that will reflect the interests of all the stakeholders	Become part of working groups to ensure their views are represented	Establish (and function as) a team to support RDM activities
Provide a champion at pro-vice chancellor research level to persuade or influence the RDM cause of the institution	Advocate RDM methods within their research communities	Analyse national, institutional and funder policy requirements
Appraise and approve proposals, plans and strategies for RDM	Collaborate with other stakeholder to gather requirements and test solutions	Identify the current status of the institution's data assets and data management practices
Advise on high-level strategic issues to be addressed during RDM service design	Support and sustain initiatives to train new researchers in good RDM practices	Develop and implement proposals, plans and budgets to develop technological and human infrastructure
Support and enable approval of an RDM policy that articulates the institutions core principles.	Voice discipline specific requirements, opportunities, challenges encountered related to RDM	Acquire/retain skills to facilitate RDM
		Advocate for RDM across the institution

		Facilitate training opportunities for managers, researchers and other support staff
		Assemble a working partnership across previously independent units

(Source: Pryor, 2014: 46-50)

To successfully rollout RDM in an institution, the three different stakeholders – university management; researchers and academic departments and support departments – will have to work together. All the stakeholders have unique roles to play in the process. Management can assemble working groups to investigate the policy of establishing RDM services in their institutions and approve the appropriate proposals or policies presented to them while also advising on high-level strategic issues. Having management buy-in might also compel researchers, who create and use data, to get involved in their institution's RDM efforts. Researchers should be involved as early as possible to voice the requirements of the disciplines. The support departments, particularly the library, IT/ICT department, and the research office will need to work together to offer and deliver RDM services.

In many HEIs, for instance Monash University (Australia) and Purdue University (USA), RDM services were initiated by the library. The next section focuses specifically on the roles and responsibilities of the library in providing RDM services.

2.5. The role of the library and LIS professionals in RDM

There is a very clear role for libraries and LIS professionals to play in supporting RDM efforts. These are described in detail in the section below.

2.5.1. The role of the library

Much of the literature on the subject of RDM identifies libraries as the initiators of RDM in academic institutions. This is because the library is considered one of the few departments in institutions that have the expertise and the capacity to offer RDM services (Walters, 2009: 85 and Heidorn, 2011: 667). Libraries have long been custodians of information objects, making them suitable to facilitate data management at their institutions (Gold, 2007: online and Cox & Pinfield, 2014: 300).

Lewis (2010) and Corrall (2012) discuss nine tactical areas in which libraries can get involved in RDM in their institutions. Each of these nine areas is referred to immediately below:

Influence national data policy

Librarians, working through their associations have the potential to influence RDM policies in their countries. US librarians influenced RDM policy through their association, the association of research libraries (ARL) (Corrall, 2012: 111). In Australia, the librarians are part of the

Australian National Data Service (ANDS) and in Canada, university library staff became members of the Research Data Strategy Working Group (Lewis, 2010: 18).

Lead-on institutional data policy

Library directors working with senior library staff members can work with administrators to understand the complexity of the challenge that is data management and develop strategies to overcome those challenges (Lewis, 2010: 18 and Corrall, 2012: 112).

Develop institutional curation capacity

In many institutions, it is the library services managing the repository. One of the challenges to managing research data is financing scalable storage, to extend on the current institutional repository to house datasets. Institutions in the UK and in the US have explored ways of developing data repositories (Lewis, 2010: 17).

Identify required data skills with LIS schools

Library managers/directors can identify data management skills and knowledge gaps and work together with LIS schools to develop a curriculum that targets these gaps (Lewis, 2010: 17). Corrall (2012: 112) adds that in the US and in the UK, libraries asked stakeholders such as funding agencies to fund LIS skills development and training to enable data management.

Develop library workforce confidence

Librarians should be responsible for keeping their knowledge up to date. Managers/directors can point them towards a wealth of appropriate resources to update their knowledge in data management. Having this knowledge will help librarians become more confident and help them engage researchers on the topic of RDM.

Bring data into undergraduate research-based learning

It makes sense that undergraduates are targeted to be taught data management skills if institutions want to improve the data management skills of postgraduates (Lewis. 2010: 14-15). Many undergraduate programmes require undergraduates to complete a dissertation, giving them experience in generating data. Data literacy can be taught to these students to equip them with good data management practices, which will prepare them for when they progress to becoming researchers (Corrall, 2012: 115).

Provide researcher data advice

Once libraries have developed the capacity (staff competencies and infrastructure) to offer data management services to their institutions, they can offer data management advice to researchers (Lewis, 2010: 13). This work can be done by extending on existing service offerings by libraries (Lewis, 2010: 13 and Corrall, 2012: 115). Libraries work with other campus service providers (IT/research computing office) to offer these services (Corrall, 2012: 116). Lewis (2010: 13) warns that at the commencement of the RDM service, the library will be limited in the advice they give to the researchers, but as the confidence of the LIS workforce increases, they will affect the way researchers approach data management.

Teach data literacy to postgraduates

Universities in the UK are involved in providing either formal or informal research training to postgraduate students. Not all these research training programmes target data management. This situation can be remedied by adding data literacy to the training programmes (Lewis, 2010: 14 and Corrall, 2012: 115).

Develop researcher data awareness

The library has a role in advocacy and creating awareness about the challenges of data management through a range of channels (Lewis, 2010: 13 and Corrall, 2012: 117). The library, working with the research office, will need to identify the disincentives for researchers and work on addressing those when advocating/creating awareness (Lewis, 2010: 14).

These nine (9) strategic areas show that libraries can get involved at various levels of RDM development. In summary: At national and institutional level, libraries can get involved by the helping develop national and institutional policies. The libraries can also get involved in building capacity. Capacity building includes developing the infrastructure and developing the confidence and skills of the employees that will potentially offer RDM services. Once the skills are developed, the employees can create awareness and offer RDM services.

Libraries have used different methods of identifying RDM services. Some used lifecycle models (CILIP, 2014; Pryor, 2014 and Tiwari & Chand, 2014) to identify areas in which libraries can play a role or offer services, whereas others aligned existing library roles with RDM roles and responsibilities (Gold, 2007; Cox *et al.,* 2012 and Charbonneau, 2013).

2.5.2. The role of LIS professionals

Working in libraries and information centres, LIS professionals have experience in working with digital information collections. The LIS professionals will be expected to carry out the service provision. In particular, subject librarians, information specialists or liaison librarians will be the professionals providing these services (Auckland, 2012; Carlson *et al.,* 2012: 215; Martin, 2012: 115; Cox & Corrall, 2013: 15 and Charbonneau, 2013: 372). These professionals are the library's first point of contact with the researchers and academic departments. The roles and responsibilities that LIS professionals are expected to take on are an extension of their current roles and responsibilities (Clements, 2013: 301 and CILIP, 2014: 8).

A report by Auckland (2012: 49) looked at the role of subject librarians in supporting the changing needs of the researchers. Managing research data, together with tasks like co-authoring research grant proposals; providing information literacy training to researchers; conducting literature reviews; writing up research methodology and offering advice on data sharing, IP rights and ownership were listed as roles that subject librarians can take on if LIS services are embedded in the research process.

Research by Charbonneau (2013: 267) found that (health science) librarians can become involved in conducting data interviews; assisting with DMPs; ensuring funder compliance by researchers; providing data literacy instruction; data sharing; publishing; citation and attribution; data impact tracking and finally curation and preservation of data.

Table 2.2 provides a list of potential RDM roles that libraries and LIS professionals can play in their institutions.

Table 2.2: Roles of the libraries and LIS professionals

Role of libraries	Role of LIS professionals
• Assisting with data management planning • Providing access to RDM platforms • Creating best practice support and guidance • Providing RDM training materials • Managing data actively • Storing data • Archiving and preserving data • Sharing and publishing data • Advising on copyright and IPR • Giving assistance to use available technology, infrastructure and tools • Providing guidance for handling and managing unpublished research data • Depositing data in an IR/subject-specific data repositories • Providing data literacy instruction • Developing tools to assist researchers with RDM • Developing an institutional policy	• Advocacy role to lead institutional data policy development • Training role (information literacy) to teach even the undergraduates about the importance of managing data • Reference services to guide researchers and provide advice about issues like copyright advice • Marketing skills to promote data re-use • Reference interview to determine who can answer certain questions • Creating a website to provide advice to researchers • Teaching metadata skills to audit to identify data sets for archiving • Collection development to develop data collections

(Sources: Gold; 2007: online; Whyte & Tedds, 2011: 4; Lyon, 2012a: 129-130; Akers & Doty, 2012: 18; Corrall, 2013: 654; Whyte, 2014: 82; Tiwari & Chand, 2014: 243; and Cox *et al.*, 2012: online).

To ensure that the LIS professionals can play a role in RDM their current skills and knowledge will have to be evaluated against the skills and knowledge required for RDM. All the developments point to the development and use of a wide variety of skills.

2.6. Knowledge and skills necessary for managing research data

The new services that libraries and LIS professionals will be offering will require that they expand on their current knowledge and skills and perhaps obtain new skills entirely. This section of the literature will look at the knowledge areas and skill sets necessary for this new role.

2.6.1. RDM knowledge areas

Knowledge of the research process/lifecycle

According to Antell *et al.* (2014: 567) and Daland (2015: 3), LIS professionals lack confidence in their ability to offer RDM support because they do not have experience and/or lack knowledge of research process. Brown *et al.* (2014: 11) suggests that RDM professionals should at least have knowledge of researcher workflows and practices. Corrall et *al.* (2013: 663) identified knowledge of the research methods and processes as one of the knowledge areas that LIS professionals need to develop.

Awareness of RDM

At the very least, librarians need to be aware of the concept of RDM (Bradbury & Brochet, 2014: 3). To further build on their RDM knowledge, individuals should have an idea about research and data lifecycles; principles and practice in data management and use; tools available for data management; metadata standards and their roles in providing RDM services (Charbonneau, 2013: 367).

Knowledge of long-term preservation of data

Charbonneau (2013: 371-372) describes data preservation as management practices based on standards that guide and build data and metadata throughout the research lifecycle and ensure its long-term care. Lyon (2007: 57) specifically mentions the responsibility of managing research data for long-term preservation of data as one of the potential RDM roles that the library/librarians can play. Knowledge of and experience in data archiving and preservation are necessary in an RDM service provider (Brown *et al.*, 2014: 11).

Knowledge of policy and governance of data

One of the main drivers (see section 2.3.2) for the management of research data is the mandates from research funding agencies (Jaguszewski & Williams, 2013: 9). Professionals

involved in offering and providing RDM services will need to be aware of the mandates and become involved in policy development, collaborating with other stakeholders such as the funding agencies and the institutions (Lyon, 2007: 57) and Brown *et al.* (2014: 11).

Knowledge of copyright and other intellectual property rights

Knowledge of intellectual property is a 'must have' skill for information professionals in general (Clements, 2013: 301). In relation to the RDM role, the RDM professionals involved must have this knowledge to be able to advise the researchers about their rights about the data they use and the data they generate (Brown *et al.*, 2014: 11).

Knowledge of metadata standards

An RDM professional should have knowledge of discipline-specific metadata standards (Charbonneau, 2013: 367 and Antell *et al.,* 2014: 567). They should be able to organise, describe research data to ensure it is easily discoverable, understood and preserved (RDA, 2015: online).

Knowledge of funders' mandates

RDM professionals will need to be aware of funder mandates to help researchers comply with these mandates (Charbonneau, 2013: 369 and Antell *et al.,* 2014: 369). Researchers are said to have poor data management skills; RDM professionals may be expected to offer researchers advice and guidance regarding data management (Charbonneau, 2013: 369).

2.6.2. RDM skills or experience

Skills or experience in indexing (metadata)

Corrall *et al.* (2013: 662) state that competency in metadata schemas is one of the key development needs for RDM support. An RDM professional will need to have metadata cataloguing skills (Gold, 2007: online; Brown *et al.* 2014: 11 and RDA, 2015: online). Ideally, the individual responsible for RDM should become involved in the research project as soon as data collection begins to record/help researchers record the appropriate metadata (Heidorn, 2011: 668).

Skills or experience online data citation/referencing

It is expected that providing access to research data will drive up citations (Brown *et al.,* 2014: 19). It is therefore important that professionals involved in RDM have necessary the skills to guide researchers on data citation and referencing (Brown *et al.,* 2014: 11). RDM professionals can support researchers in this regard by offering advice on data citation, assigning identifiers to datasets, linking research output to datasets and helping (re-)users attribute credit to data creators (RDA, 2015: online).

Skills or experience in data management planning

Individuals involved in offering RDM services will need to provide guidance on data management plans to researchers who are applying for research funding (Corrall, 2013: Brown *et al.,* 2014: 11 and Davidson *et al.,* 2014: 217). These individuals will need to have an understanding of the various research lifecycles, workflows and CoPs to be able to advise researchers on a variety of data management planning issues (Charbonneau, 2013: 368).

Skills or experience in creating guides or training material for researchers

Brown *et al.* (2014: 11) identify training materials as a necessary skill to have by an RDM professional. Charbonneau (2013: 370) sees the need for data management as an opportunity for libraries to expand on data instruction.

Skills or experience in organising and documenting data

Data generated from research exists in different formats. Quality control then becomes an important part of the research process. This role is often left to the researcher. The role of the RDM professional will be help the researcher to organise data files in a logical manner to facilitate retrieval and re-use of data (Corti *et al.,* 2014: 67). Lavoie (2012: 67) stresses the importance of providing sustainable RDM services to ensure sustainable access to research data.

Skills or experience in data storage and security

A good strategy is required for securely storing, backing up, transmitting and possibly disposing of research data (Corti *et al.,* 2014: 86). It is not anticipated that the LIS workforce will be expected to actively store and ensure security of the data. They can play an advisory role, however, for researchers and work together with the IT department and the research office.

Skills or experience in data licencing

The individuals responsible for RDM will be expected to help researchers share the data they created using the relevant licences (RDA, 2015: online). It would be beneficial for those responsible to have some knowledge/experience with data licencing. Data citation standards vary across disciplines; however, there are tools such as DataCite that seek to make citing data easier. The individual responsible for RDM should be aware of such tools and be able to assist researchers in using them (Charbonneau, 2013: 370).

Skills or experience in working with and managing repositories

It is expected that datasets will be stored in repositories, whether data repositories or institutional repositories. The RDM workers will need to have worked with repositories whose aim is to publish and archive research data/datasets (RDA, 2015: online). Heidorn (2011: 666) suggests special expertise is needed in running institutional repositories.

Skills or experience in data appraisal and selection

Not all data that is generated will need to be retained beyond the lifecycle of the project (Ashley, 2012: 156). Gold (2007: online) mentions data appraisal and selection as one of the services that the libraries can provide in enabling and supporting the management of research data. This means the professionals offering RDM services will need to have skills in appraisal and selection of data (Heidorn, 2011:668).

It is not expected that one individual should possess all the RDM skills and knowledge discussed above. The combination of RDM skills and knowledge an RDM professional need to possess will vary according to their institution and the services the institution offers. The knowledge and skills discussed above exclude the soft skills that LIS professionals should already have. Examples of such skills are: are communication skills; computer literacy skills; networking skills and the ability to manage or supervise people (Pryor & Donnely, 2009: 165; Antell *et al.,* 2014: 567 and Pryor, 2014: 53).

LIS professionals already possess some of the RDM skills and knowledge because they are aligned with their existing roles (Corrall *et al.,* 2013: and Cox *et al.,* 2012: online). However, translating their existing skills and knowledge into RDM is a challenge (CILIP, 2014: 9 and Jones, 2014: 6).

Owing to the ever-changing landscape of technology and subsequently e-Research, which includes e-Science and Digital Humanities, LIS professionals cannot become complacent.

Where necessary, they will have to enrol for appropriate courses to improve the knowledge and skills they may be lacking (Corrall *et al.,* 2013: 646 and Antell *et al.*, 2014: 567).

2.7. Training opportunities available

There are numerous RDM-related training opportunities available to the LIS workforce. Unfortunately, the training that the LIS professionals attend is responsive rather than opportunistic/proactive (Auckland, 2012: 60).

Libraries' response to the shortage of appropriately skilled professionals could be to recruit new personnel who are already experts in the field of RDM. Alternatively, libraries can choose to re-skill their existing workforce (Pryor, 2014: 53). A number of RDM training opportunities have been designed specifically for re-skilling LIS professionals (Lyon, 2012a: 133 and Corrall *et al.,* 2013: 647). Examples of such programmes are provided in section 2.7.2 below. Some HEIs are addressing the skills shortage at graduate level (see section 2.7.1).

Heidorn (2011: 671) recorded two examples from Illinois and Arizona. The University of Illinois at Urbana Champagne offers a Data Education Program. University of Arizona offers a course in digital collections that focuses on digital formats, metadata and data migration, applied technologies, digital collections management, preservation and institutional organisation (Heidorn, 2011: 671). South African options are provided in the next section.

2.7.1. University-level training

LIS professionals have the option to enrol in formal university degrees. It makes sense for early career LIS professionals to consider these options, especially undergraduate degrees. The university-level training opportunities have been slower to develop than in-house training opportunities. The degrees that are available have been established, differ in degree/diploma type and level. The one is an undergraduate degree and the other is a postgraduate degree. The master's level degree has a much narrower focus compared to the BSc degree. Both of these provide a good background to the graduate so that RDM could be better supported.

a. Bachelor of Science (BSc) Data Science degree

The Sol Plaatjie University (SPU) became the first university in South Africa to offer an undergraduate degree in data science. This is a three-year undergraduate degree, which seeks to equip students with the competencies necessary to solve big data issues (SPU, 2015: 4). This programme is not specifically aimed at practicing LIS professionals, rather individuals who are interested in of data science – not only supporting or facilitating data management.

b. MPhil Data Curation

At the University of Cape Town (UCT) in South Africa, the school of library and information services offers a Master's degree in digital curation. The curriculum addresses the following:

- Information architecture and metadata;
- Technology platforms
- RDM
- Digital curation principles, theory and philosophy

(UCT, 2015a: online)

The duration of the degree is two years. The individuals who choose to enrol in the programme will be expected to complete the coursework (curriculum) as well as a mini-dissertation on a topic related to data management (UCT, 2015a: online).

2.7.2. In-house training interventions at HEIs

For many LIS professionals it would make more sense to enrol in short courses instead of completing a full degree course. The DCC and JISC have designed and piloted many training programmes for LIS professionals but several other institutions have done similar work. Most of these are available online for individuals or institutions to use to learn more about RDM.

a. New England Collaborative Data Management Curriculum (NECDMC)

This training programme was designed by Lamar Soutter Library at the University of Massachusetts to train their (science) librarians in data management. This training programme was then adapted to train science, technology, engineering and math (STEM) education librarians at the University of Manitoba in Canada (Kafel *et al.,* 2012: 62 and Ishida, 2014: 82). This programme was built on the Framework for Data Management

Curriculum designed to train the undergraduate and postgraduate students in STEM fields (Lamar Soutter Library, 2015: online).

The programme consists of seven (7) modules that take five days to present over the course of four weeks. The modules are:

- Research data management overview;
- Data types, formats, and stages;
- Metadata;
- Data Storage, backup and security;
- Legal and Ethical Considerations;
- Data sharing and reuse; and
- Repositories, archiving and preservation.

The information provided above was reported by Kafel *et al,* (2014: 63) and Ishida (2014: 82).

A train-the-trainer programme was designed together with the NECDMC curriculum. The train-the-trainer programme is aimed at teaching the trainers in instructing the curriculum (Kafel *et al.,* 2014: 64).

b. **SupportDM/TraD – DCC and University of East London**

Training for Data Management (TraD) is a JISC-funded training programme developed for the librarians at the University of East London (UEL). Guy (2013: online) notes that the training was based on the MANTRA training developed by the University of Edinburgh. The training curriculum consists of five (5) modules namely:

- About research data management;
- Guidance and support for research;
- Data management planning;
- What data to keep, and why; and
- Cataloguing data

(University of East London, 2015: online).

c. Digital Curation (DC) 101: How To Manage Research Data

DC 101 was an introductory workshop to RDM and curation. This particular workshop is being offered free of charge, conducted over half a day (approximately 4 hours). The workshop covers the drivers and benefits of RDM and the range of activities and roles to consider when planning new projects (DCC, 2015: online).

This training is not directed at LIS professionals in particular. However, it would be beneficial for them to have some knowledge of what is expected of the researchers to better support them.

d. RDMRose

RDMRose is continuous professional development (CPD) type training programme funded by JISC (University of Sheffield , 2015: online). Being a collaboration project by Leeds, Sheffield and York University libraries, the training programme was developed to train librarians and other information professionals to support RDM. The module has eight themes, namely:

- Introduction and role of LIS in RDM;
- Nature of research and the need for RDM;
- The digital curation lifecycle;
- Key institutions and projects in RDM;
- What is data?
- Managing data;
- Case studies of research projects; and
- Institutional case study and conclusion

University of Sheffield (2015: online)

The themes are broken down into four (4) parts, presented over four (4) hours or half a day University of Sheffield, 2015: online and DCC, 2015: online).

e. RDMRose Lite

RDMRose Lite, based on the RDMRose training described above, is a three-hour introductory RDM course for librarians (RDMRose, 2015: online). It was developed in collaboration with the DCC and the University of Northampton. The course is much shorter in duration than the RDMRose programme and covers:

- Research data and RDM;
- Data management planning;

- Data sharing; and
- RDM skills

(RDMRose, 2015: online and DCC, 2015: online).

f. DIY RDM Training Kit for Librarians

The DIY RDM Training Kit for Librarians developed by EDINA and the Data Library at the University of Edinburgh is a training programme aimed at helping information professionals understand core issues related to RDM and their role in supporting RDM at their institutions (Jones, 2014: 2 and Rice, 2014: 49).

The training covers five modules including:

- DMP;
- Organising and documenting data;
- Data storage and security;
- Ethics and copyright; and
- Data sharing

(Rice, 2014: 49)

Each session is two (2) hours in duration and the training is conducted face to face, facilitated by data librarians and/or guest speakers. This training provides practical understanding of core methods and tools (DCC, 2015: online).

The number of training opportunities available to LIS professionals is steadily increasing as more institutions are planning for and implementing RDM services. The training material that has been made available by the DCC and JISC in the UK has allowed many universities to develop their own specialised in-house training programmes to train their workforce.

Although the in-house training programmes discussed above range in duration and number of modules offered, much of the content is similar. Modules that are prominent in the training opportunities discussed above are:

(1) RDM overview/introduction,
(2) DMP,
(3) The role that the library and LIS is (expected) to play in RDM,
(4) Data sharing and re-use.

The DIY RDM Training Kit for Librarians curriculum was used to design many of the training programmes, which could be considered as the reason for the similarity in the training opportunities. These modules all address the knowledge component of RDM training but they do not really develop actual skills.

There are modules unique to each training opportunity. RDMRose for example covered key "Institutions and projects in RDM"; "case studies of research projects" and "institutional case study and conclusion". NECDMC's most recent module covered "Repositories, archiving, and preservation while the most recent module in SupportDM/TraD covers "cataloguing data". These options may be seen to at least start addressing the skills component of RDM competency development.

2.8. New positions for RDM

Proctor *et al.* (2012: 142) points out that HEIs and their libraries are expected to reduce their (operational) costs, meaning that most of the libraries cannot afford to hire new personnel. This shortage of skills may hinder the library's progress in establishing RDM services. As a solution, libraries may choose to re-skill their existing LIS workforce, particularly science librarians, to take on the RDM role (Charbonneau, 2013: 366 Antell *et al.,* 2014: 570).

Auckland (2012: 53) mentions that some job descriptions for "regular" subject librarians now expect candidates to contribute to emerging services such as supporting RDM. It would makes sense that new job titles/positions are created for newly recruited RDM experts or that re-skilled LIS professionals take on the RDM role. Lyon (2012a: 131) described the roles and responsibilities of data professionals (data scientists, data librarians, liaison librarians or faculty librarians). Below are some examples of job positions created for professionals that will be offering RDM services.

a. Data scientist

In recent times, the position of "data scientist" has gained popularity, with Davenport and Patil (2012: 72), calling this the sexiest job of the 21st century. These professionals will be responsible for delivering data advice and guidance to the staff. These individuals will also be responsible for facilitating access to datasets for researchers. They will need to have discipline-specific skills and knowledge, together with informatics skills, to provide effective data management services. There is a shortage of individuals with these skillsets and knowledge (Davenport & Patil, 2012: 72). Data scientists mostly prefer to work in subject-

specific research institutions, research councils, and in research projects and they least prefer working in computer centres, data support services and large data centres (Swan & Brown, 2008: 17). It was also established that most data scientists become data scientists by serendipity (Swan & Brown, 2008: 1) rather than being specifically trained for the position.

b. Data manager

DCC (2011: 2) describes data managers as managers and curators of research data generated from research. Typically, data managers check and clean data, prepare documentation about datasets, handle metadata, provide access to datasets to users, prepare data for analysis and archive data. The data management profession is still new, therefore, there are still very few formal curricula available to develop this profession. In most HEIs, data managers have experience in fields such as IT, computer science, information science and library science (DCC, 2011: 2).

c. Research data librarian

At the University of Maryland, two Research Data Librarians (RDLs) positions have been filled. The RDLs offer guidance, project consultation and technical assistance on aspects of data curation and management. Consultation services on DMPs, data sharing and publication, and long-term preservation are offered (University Libraries, 2015: online).

d. Informationist

Gore (2013: 20) worked as part of a research team at the University of Massachusetts Medical School and was given the title “informationist”. As part of the job responsibilities, Gore had to develop data management tools, provide a detailed literature review, report on issues facing researchers and IT specialists when developing or implementing research tools (Gore, 2013: 20).

In SA, it was possible to retrieve four positions that were advertised. These are provided here as a continuation of the list of possible job titles.

e. Research data curator

At the University of South Africa (UNISA), the minimum requirements for the above-mentioned position is a three-year LIS degree and five years’ work experience (UNISA, 2015: online). The skills required for a researcher data curator at this institution are interpersonal-,

communication- and collaboration skills, as well as knowledge of research methodology and research data. Knowledge of risk assessment and management and knowledge of metadata and data-curation standards are required (UNISA, 2015: online).

f. Senior data scientist

For this position, the prospective candidates are usually required to have a PhD (doctoral degree) in computer science or any other IT-related degree and work experience (UCT, 2015b: online). Knowledge or experience in the following areas is also required: structured or non-structured database programming, web application technologies, scripting languages, and the ability to deploy and troubleshoot third-party applications (UCT, 2015b: online).

g. Digital curator

At UCT, the position of digital curator requires the candidates to have a three-year LIS degree and two years' experience (UCT, 2015c: online). The prospective digital curator will need to have: knowledge of existing and emerging technologies, standards, and best practice in digital repository discovery and preservation activities; skills in curation software and tools; metadata standards; information management problem solving; and collaboration skills (UCT, 2015c: online).

The main responsibilities associated with this job are digital data management, digital curation activities, digital repository management, resource management and project management (UCT, 2015c: online).

h. Assistant director: Research data management

At the University of Pretoria (UP), the position of the person responsible for RDM is "Assistant Director: Research Data Management". This professional's key responsibilities are to provide RDM guidance to faculty, provide input on RDM to UP executive level, develop RDM policy and guidelines, and to establish RDM pilot projects with faculty (UP, 2015: online).

HEIs have recently begun planning for and implementing RDM services. The number of job positions that have already been developed shows progress. With these jobs, the professionals are expected to perform the tasks similar to those that were discussed in section 2.5. These job tasks include offering RDM advice to researchers, actively managing data and making past data available to researchers.

The breadth of competencies (skills, knowledge, experiences and personality/behavioural traits) required to enable RDM vary by institutions and the types of services they offer. Competencies that are seen across the various job positions/titles are (1) subject knowledge, (2) qualification(s) or experience in library- or information science, and – in some instances – (3) IT or computer science.

Subject knowledge is useful as the professionals are sometimes expected to become involved in research projects. The qualifications in the above-mentioned areas could help professionals actively manage data. A background in IT should help when required to develop tools or applications to facilitate the research process and train researchers to use these tools.

The training opportunities available (discussed in section 2.7.) can help LIS professionals gain the knowledge and skills required. LIS professionals have experience in working with digital materials and working with researchers. The added training could logically make them the most competent professionals to take on the RDM role.

2.9. Challenges to initiating RDM by LIS professionals

Section 2.3. argues the need and urgency for implementing RDM services in HEIs. Yet, issues exist that may hinder efforts to initiate RDM. These issues are discussed below.

Lack of personal experience with research

Librarians are said to lack personal experience with the research process (CILIP, 2014: 9 and Daland, 2015: 2). Antell *et al.* (2014: 566) identified knowledge of the research process as one of the areas of importance in enabling RDM. The individuals/service unit responsible for providing an RDM service provider will need to have knowledge of the research process in a specific domain. Such a provider might be rare to find.

Lack of domain-specific knowledge

Research lifecycles/process differ in different disciplines/domains (Cox *et al.,* 2012: online). Therefore, general knowledge of the research process may not be enough to support or facilitate RDM in HEIs. Knowledge of the subject/discipline is very important for managing research data (Auckland, 2012: 37 and Antell *et al.,* 2014: 596). Cox *et al.* (2012: online) and Jones (2014: 6) listed the librarians' lack of domain knowledge as one of the challenges to RDM.

Translating current practices is not easy

Librarian's jobs entail expertise that is thought to overlap with RDM. Theoretically, librarians have the expertise required for RDM (CILIP, 2014: 9 and Cox & Pinfield, 2014: 300). Cox *et al.* (2012: online) argue that translating current library expertise and practice to RDM is not as straightforward as it appears. The technical knowledge required for RDM is linked to records management and archiving more than it is linked to library science (CILIP, 2014: 9).

Time constraints

Should the library take on RDM services, it would be difficult to find the time to offer such services as librarians are already over-stretched (Martin, 2012: 115 and Cox & Pinfield, 2014: 300). Cox *et al.* (2012: online) state that librarians will have to cast aside some of their tasks/responsibilities for taking on this new role.

Inability to engage researchers

Librarian's inability to engage researchers with LIS services is seen as a challenge to RDM (Cox *et al.,* 2012: online; CILIP. 2014: 9 and Jones, 2014). Coates (2014: 56) offers a different opinion, stating that libraries are members of the institutional research support community with relationships and social interactions with researchers. Coates (2014: 56) suggests that librarians' personal relationships with researchers will facilitate researcher engagement.

Other non-skills related challenges to RDM

Lack of (formal) policy, infrastructural issues, data legacy are some of the most common issues that hinder data management (Jones *et al.,* 2008: 116-118; Cox *et al.,* 2012: online and CILIP, 2014: 9).

The majority of the challenges that HEIs and their libraries face in offering RDM services are knowledge- and skills related. Although the LIS professionals are the most competent professionals to take on RDM in their institutions, further skills development is needed. Here 'skills development' not only includes the RDM skills and knowledge. Soft skills such as communication and negotiation skills should also be considered. In addition to the skills challenges, lack of policy, data legacy issues and underdeveloped infrastructure are also important factors to consider. The challenges should not be addressed individually; all the relevant stakeholders should work to resolve all of the challenges to ensure true progress.

2.10. Conclusion

Managing research data has become a priority for in HEIs because of the national and institutional understanding regarding the value of research data. Although there are numerous stakeholders involved in enabling and supporting RDM, it is the library and LIS professionals who are expected to lead this initiative to take care of and make data accessible in the longer term. Reason being that libraries are one of the few departments with the capacity to work with digital information objects and LIS professionals have experience in working with/managing digital information objects. This makes these professionals the most suitable candidates for RDM.

The DCC and JISC have made RDM training material freely available to the public. Libraries have the option of making use of the training material to develop their own in-house RDM training for LIS professionals. Alternatively, interested professionals have the option of enrolling for formal university degrees, although these options are few.

There is a large range of skills that need to be developed if managing research data is to be carried out correctly. The skills will in all probability only become clearly defined once the various roles in managing data are defined. At this early stage any professional wanting to work with research data would need to at least have a clear understanding of the research process and the data cycle. It is also necessary to acknowledge differences in the data collected for and used in various disciplines. Only then could the hard- and soft skills necessary to actively manage any research data be developed.

CHAPTER 3: RESEARCH METHODOLOGY

3.1. Introduction

The purpose of this chapter is to describe the research methodology used to conduct the investigation of whether a specific RDM-training event might upgrade LIS professional's RDM understanding, knowledge and skills. This chapter covers the research paradigm, the research design, the data-collection methods and the tools used to conduct the investigation. The chapter also describes the research population and the pilot study conducted. Finally, a brief discussion of the data analysis will be provided before the chapter is concluded.

3.2. Overview of the research

The research was conducted to investigate if a specific training event could improve the RDM knowledge and skills of the LIS professionals in HEIs. Research sub-questions were formulated to help answer the central research question.

The central research question

The research sub-questions are listed below.

i. What is seen internationally as RDM?
ii. Should libraries and LIS professionals play a role in their HEI's RDM efforts? What could these roles entail?
iii. What is the best way to obtain or improve knowledge and skills of RDM?
iv. To what extent would a structured RDM training intervention enhance the RDM understanding, knowledge and skills of LIS professionals at SA HEIs?

To answer these questions, the researcher considered various ways to conduct the research. Section 3.3 discusses the research paradigm and section 3.4 the research design used to conduct the research.

3.3. Research paradigm

Quantitative research and qualitative research are seen as the two broad research paradigms available to an empirical researcher. Donmoyer (2008: 716) describes quantitative research as approaches to empirical inquiry that collect, analyse, and display data in

numerical rather than narrative form. Dawson (2009: 15) adds that quantitative research generates statistics through the use of large-scale-survey research, making use of data-collection methods such as questionnaires or structured interviews.

Qualitative research is viewed as an umbrella term used to cover a wide variety of research methods and methodologies that provide holistic and in-depth accounts to reflect complex nature of our social world (Staller, 2010: 1159). Under the qualitative research paradigm, attitudes, behaviours and experiences are explored through the use of data collection methods such as interviews or focus groups (Dawson, 2009: 14).

A third paradigm, called “mixed methods research”, may be considered. Mixed methods research combines at least one qualitative component and one quantitative component in a single research project (Bergman, 2008: 1 and Creswell, 2014: 217).

This study is considered to be a qualitative study rather than a quantitative study because the study sought to investigate the extent to which a particular program could enhance the understanding, knowledge and skills of LIS professionals.

3.4. Research design

A research design is the plan, structure, and strategy of investigation developed to help achieve research objectives and answer the research questions (Kumar, 2011: 94). A number of research designs are available to researchers. These include case study research, grounded theory research, phenomenology, and experimental and quasi-experimental research designs. A researcher can decide to use more than one research design called an “embedded research design” (Edmonds & Kennedy, 2012: 158).

.

The researcher used an embedded research design, combining a quasi-experimental design which is a quantitative research design and a case study research design, a qualitative research design. Sections 3.4.1. discusses the case study research design and section 3.4.2 the quasi-experimental research design.

3.4.1. Case study research

Pickard (2013: 101) suggests that the term “case study” has perhaps been used too broadly, which had led to the term being used to define research that does not fit into other research design categories. However, the case study research design is a qualitative research design

that, according to Litchman (2014: 118), examines a particular person, programme, event or concept. Researchers can focus on a particular case because it is unique and can provide better understanding of the case or inform practice(s). Researchers can also choose to study two or more cases that are different in some ways for making comparisons, building theory or even proposing generalisations (Leedy & Ormrod, 2010: 137).

There are three types of case studies: intrinsic case studies, instrumental case studies and collective case studies (Creswell, 2013: 99 and Pickard, 2013: 102).

- The *intrinsic case study* provides a better understanding of the case. The case is studied for its ordinariness as much as it is for its peculiarities.
- The i*nstrumental case study* examines a phenomenon or theory and the case becomes less important other than as a vehicle for the investigation.
- The *collective case study* studies multiple case studies to investigate particular phenomena.

(Creswell, 2013: 99 and Pickard, 2013: 102)

This study can be seen as an intrinsic case study. The intention of the study was to help the researcher better understand a single case – a specific RDM workshop programme as a vehicle to increase, knowledge, understanding and skills. This was a unique study, the first of its kind in SA.

3.4.2. Experimental research

Experimental research, in a quantitative research design, attempts to establish causality; that is, the dependent variable is a direct result of the independent variable. In true experimental research, the researcher manipulates the independent variable and examines the effects on the dependant variable (Leedy & Ormrod, 2010: 228). In experimental research, the independent variable is the phenomenon/variable manipulated by the researcher, whereas the dependent variable is the behaviour/effect measured by the researcher (Pickard, 2013: 120). The three broad categories of experimental design are the pre-experimental design, the true experiment and the quasi-experiment (Leedy & Ormrod, 2010: 228; Kumar, 2011: 113 and Pickard, 2013: 121).

The pre-experiment

In the pre-experimental research design, also referred to as the non-experimental design (by Kumar, 2011: 114), it is not possible to show the cause and effect relationships because

either the independent variable doesn't vary or the control group does not comprise randomly selected individuals (Leedy & Ormrod, 2010: 229).

The true experiment

In a true experiment, the researcher starts from the cause and seeks to determine the effects. The independent variable can be introduced, observed, controlled and manipulated by the researcher. The reason for this is to empirically verify a hypothesis of a causal relationship between variables (Kumar, 2011: 114).

The quasi-experiment

Pickard (2013: 123) states that the quasi-experiment differs from the true-experiment in purpose and process. While the true experiment aims toward covariance the quasi-experiment aims to establish levels of correlation between observable variables (Pickard, 2013: 123). Quasi-experimental designs are used when it is not possible or practical to select a random sample. Researchers may also use the quasi-experimental design if alternative explanations cannot be ruled out (Leedy & Ormrod, 2013: 237 and Kirk, 2013: 7).

The experimental design that was most suited for this research was the quasi-experimental design. As both the case study and the experiment were used to conduct the research, the research design is therefore identified as an described as the "embedded design". The embedded research design is discussed in section 3.4.3.

3.4.3. The embedded design

The case study design and the quasi-experimental design were used to conduct this study. The quasi-experimental design type used for study was the non-randomised control group pre-test-post-test design (Leedy & Ormrod, 2013: 237). The group used for observation was naturally occurring. Group members were invited to attend an RDM workshop organised by the NeDICC.

The quasi-experimental research design was designed around the RDM workshop hosted by NeDICC in August 2015. The workshop became the training intervention. Two online questionnaires were designed to measure the LIS professional's (participants) perceptions of their RDM understanding, knowledge and skills before and after the RDM workshop (training intervention). The participant's answers to the before and after questionnaire were then compared to determine whether the training intervention changed the participant's perceptions of their RDM understanding, knowledge and skills.

Figure 3.1 depicts the research design used to conduct the research:

Group = O1 X O2

Figure 3.1: Quasi-experimental design

The symbols used for this experiment are the following:

O = observation/measurement

Observation refers to the online questionnaires the participants had to complete pre-test (O1) and post-test (O2).

X = Intervention

Intervention refers to the training intervention that the participants attended: the RDM workshop.

Group = the group of participants who completed the questionnaires.

Figure 3.2 illustrates the sequence of the experiment. In the first block, the most significant results attained from the first questionnaire are shown. In the second block the contents of the intervention are displayed. In the third block, the most significant changes resulting from the intervention are shown. Section 3.5 discusses the data-collection methods available to an empirical researcher.

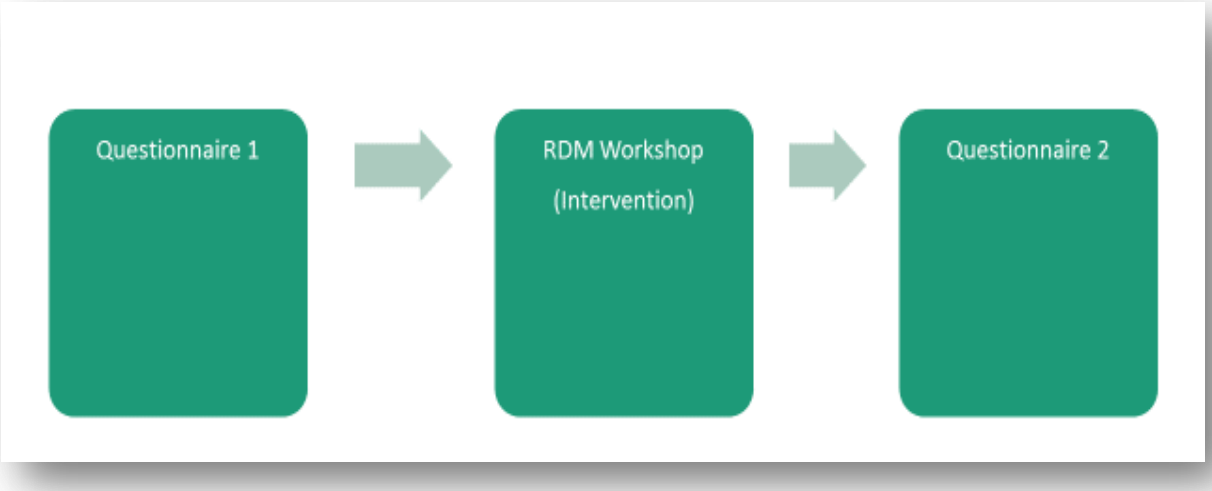


Figure 3.2: Quasi-experimental model used for this study

3.5. Data-collection methods

Chapter 2 helped the researcher identify the data-collection methods and tools used to conduct similar studies in the past. In section 3.5.1, the different data collection methods the researcher considered to conduct the research are discussed.

3.5.1. Observation method

Kumar (2011: 140) states that observation is a method of collecting primary data. Observation is a purposeful, systematic, and selective way of watching and listening to a phenomenon as it takes place (Kumar, 2011: 140). Observation under the qualitative research paradigm is often unstructured and free flowing, allowing the observer to switch focus from one thing to another (Leedy & Ormrod, 2010: 182).

Observation as a data-collection method has the following advantages:

- Observation offers flexibility;
- Observation is used when the research participants cannot provide objective information; and
- Observation can also occur within other data-collection methods such as interviews.

(Leedy & Ormrod, 2010: 182 and Kumar, 2011: 140)

The disadvantages of observation as a data-collection method are the following:

- Novice researchers will not always know what things are most important to look at/observe;
- Researchers can waste time;
- Research participants being observed may change their behaviour if they become aware of the fact that they are being observed;
- The possibility of incomplete observation exists;
- The possibility of observer bias exists; and
- Interpretation of the observation may differ according to the observer

(Leedy & Ormrod, 2010: 183 and Kumar, 2011: 141)

3.5.2. Interviews

One of the most common methods of collecting data from people is through interviews (Kumar, 2011: 144). Research interviews can take numerous forms – formal and structured

to informal and purposeful conversation (Pickard, 2013: 195). Leedy & Ormrod (2010: 188) state that interviewing in qualitative research studies tends to be informal and friendly whereas interviews in quantitative research are more structured and formal. Interviews are used when the researcher seeks qualitative, descriptive and in-depth data that is specific to the individual and when the nature of data is too complicated to be asked and answered easily (Pickard, 2013: 196).

Structured interviews are known for their rigidity in structure, interview content and wording of interview questions. Unstructured interviews are used for their flexibility in structured, interview contents and questions (Kumar, 2011: 144). Structured interviewing can take place in two forms which are *standardised, open-ended interviews* (all interviewees are asked the same questions and are allowed to answer anyway they please) and closed fixed response interviews (interviewees are asked the same question and choose from a predetermined set of alternative answers) (Pickard, 2013: 199).

Unstructured interviewing is used to gain a holistic understanding of the thoughts and feelings of the interviewee (Pickard, 2013: 199). Unstructured interviews can be conversational interviews (purposeful conversation where questions are allowed to flow naturally). Alternatively, a guided interview can be conducted where a researcher prepares a checklist to ensure all relevant areas of a topic are covered during the interview (Pickard, 2013: 200).

Advantages of interviews as a data-collection method

- Interviews are useful for collecting in-depth information;
- It is possible to supplement information;
- Questions can be explained and ambiguity clarified;
- Application is wider; and
- The return rate is high

(Kumar, 2011: 150, Leedy & Ormrod, 2013: 191 and Pickard, 2013: 207).

Disadvantages of interviews as a data-collection method

- Interviews can be time consuming and expensive to conduct;
- Quality of data collected depends on the quality of the interaction;
- Quality of data collected depends on the quality of the interviewer;
- Data quality may vary if different interviewers are used; and

- The researcher may introduce his or her own bias during the interview

(Kumar, 2011: 150 and Leedy & Ormrod, 20133: 191)

3.5.3. Questionnaires

Kumar (2011: 145) describes a questionnaire as a written list of questions given to participants to read, interpret and answer. The difference between the questionnaire and structured interviews is that in structured interviews the researcher asks and if necessary, explains the questionnaire to the participant(s) (Kumar, 2011: 145 and Pickard, 2013: 207). Pickard (2013: 222) identified three methods of collected data through questionnaires: paper-based; electronic and researcher-administered questionnaires. The paper-based questionnaires, which are the most common types of questionnaires are printed and distributed in a variety of ways including post, hand-delivery, or leaving them in a position where participants will be encouraged to complete and return the questionnaires (Pickard, 2013: 222).

The electronic questionnaires refer to questionnaires that are either online or sent within an email message (expect questionnaires sent as an attachment). It is located on a web-page where participants can complete the questionnaire online (Pickard, 2013: 222).

Advantages of questionnaires as data-collection methods

- Administering questionnaires is often less expensive than the other methods discussed above;
- Questionnaires offers great anonymity compared to other methods;
- Questionnaires can be distributed to a larger sample size;
- Data analysis can be determined from the onset; and
- Participants may answer more truthfully if the questionnaire is anonymous.

(Leedy & Ormrod, 2010: 189 and Kumar, 2011: 148)

Disadvantages of questionnaires as data-collection methods

- Application is often limited to a population that can read and write;
- Questionnaires have a risk of low response rate;
- Self-selecting bias;
- There are no opportunities to clarify ambiguity to the participants;
- No opportunity for spontaneous questions; and

- Response to questions may be influenced by the response to other questions.

(Leedy & Ormrod, 2010: 189 and Kumar, 2011: 149)

Online questionnaires were used as the data collection method for this study. The primary reason for using online questionnaires was to reach more participants. The online questionnaires were anonymous, participants answer more truthfully in such situations.

3.6. Data collection tools

Two online questionnaires were designed to collect data for the study. These questionnaires were similar except for questions about the presentations and value of the workshop added to the second questionnaire. The questionnaires are located in the appendix section, labelled as Appendix A (Questionnaire 1) and Appendix B (Questionnaire 2).

The questionnaires comprised open- and close-ended questions. Open-ended questions were asked to get the personal opinions of the participants. Open-ended questions bring out descriptive data, which added detail to the close-ended questions (Pickard, 2013: 210-211). Close-ended questions asked in the questionnaire were dichotomous and multi-dichotomous questions. This means that the participants could choose an answer from a list of options (dichotomous) or multiple answers from a list of options (multi dichotomous) (Pickard, 2013: 211). Section 7 discusses the research population.

3.7. Research population and location

Pickard (2013: 60) defines research population as "…*the entire set of individuals about which inference will be made*". The population of this study was the attendees of the RDM workshop. It was expected that 65 people would attend the workshop as they were required to register and pay before they could attend the event. No sampling method was used because the entire research population was asked to participate in the study.

The training intervention took place at the Knowledge Commons of the Council for Scientific and Industrial Research (CSIR). The CSIR is located in Pretoria, Gauteng. The participants completed the online questionnaires at their varying locations.

3.8. Pilot study

A pilot study was conducted prior to the main investigation to determine feasibility and to identify potential deficiencies of the questionnaires. A group of five LIS professionals - two lecturers from the Department of Information Science at the University of Pretoria and three information specialists at the Department of Library Services at the University of Pretoria were asked to complete the online questionnaires. Three of the five individuals replied. The general response was that the questions were easy to answer and not time consuming. One pilot participant suggested that the researcher should add a message at the end of the first questionnaire to remind the participants that they would be asked to complete a second questionnaire after the workshop. This suggestion was incorporated into the questionnaire.

3.9. Data analysis

Data for the research was collected using online questionnaires. The questionnaires were designed using the web-based application Google Forms. The programme captures the participant's answers and saves them so that the researcher can access them when necessary. The data collected can be accessed in a variety of formats such as Excel spreadsheet (xlsx), pdf document, open document (ods), comma separated values (csv) and tab separated values (tsv). The researcher thought the data would be easiest to analyse in the xlsx format.

The questions asked were displayed in rows (horizontally) and all the participants' answers were displayed in the columns (vertically). The researcher was then able to consider the participant's answers collectively for each of the questions in the questionnaires (Appendix 1 and 2). The data collected were analysed through inductive reasoning, sorting and categorising to identify themes (Leedy & Ormrod, 2013: 158).

3.10. In summary

This chapter explored the research methodology used to conduct the research. The research paradigms were described. It was revealed that the research was conducted under the qualitative research paradigm (section 3.3). An embedded research design was used to guide the research to help answer the research question and sub-questions. The quasi-experimental design used is depicted in figure 3.1. and figure 3.2.

Data was collected using self-administered online questionnaires designed using Google Forms. The research population was a number of LIS professionals attending the RDM workshop organised by NeDICC. A small pilot study was conducted to test the feasibility of the questionnaires before implementing the larger study.

Data analysis and findings are discussed in the next chapter.

CHAPTER 4: DATA ANALYSIS

4.1. Introduction

Chapter 3 discussed the research methodology used to conduct the research. This chapter presents findings from the research data collected using online questionnaires guided by an embedded research design. The next section will describe how the data was collected and the intervention applied.

4.1.1. Background to data collection

The NeDICC hosted an RDM workshop at the CSIR on the 11th of August 2015. Library and Information Science (LIS) professionals from SA Higher Education Institutions (HEIs) and research organisations were invited to register for the RDM workshop. Although the focus of this research is on HEIs the researcher did not exclude the responses from LIS professionals coming from research organisations because their RDM challenges are very similar.
Sixty-three (63) people attended the workshop, four (4) of whom were presenters and one the researcher. This then makes the total population 58.

On the 5th of August, a week before the workshop took place, an email was sent to the individuals that had registered for the workshop, asking them to participate in the research and providing them with link to the first questionnaire. A reminder was sent to the individuals on the 7th of August, asking them to complete the questionnaire if they hadn't done so previously. On the 11th of August, the questionnaire was closed off to the public and the researcher had received a total of thirty-one (31) responses. Upon closer inspection, the researcher discovered that one of the presenters had completed the questionnaire and the decision was made to remove the entry. This made the number of responses 30 instead – a return rate of 52%.

On the 12th of August an email was sent to the attendees of the workshop the link to the second questionnaire. And two (2) reminders were sent on the 13th and the 28th of August. The questionnaire was closed to the public on the 31st of August. Table 4.2 summarises the data-collection process.

Table 4.1: Background to data collection

Date (2015)	Activity	Response rate
First questionnaire (Observation 1)		
5 August - Request to complete questionnaire was sent 7 August – Reminder was sent to attendees to complete the questionnaire 11 August (08:00) - questionnaire closed to public	Knowledge and skills required for research data management – before training intervention!	31 responses to the questionnaire 30 responses after removing one entry 52% response rate
Intervention (NeDICC workshop on RDM)		
11 August	RDM Workshop attended by NeDICC (X) Intervention	58 attendees
Second questionnaire (Observation 2)		
12 August – Request to complete 2nd questionnaire sent 13 August – Reminder to complete questionnaire sent 28 August – Final reminder to complete questionnaire sent 31 August – questionnaire closed to public	Knowledge and skills required for research data management – after training intervention! (Observation 2)	21 responses to the questionnaire. 36% response rate

The responses from both questionnaires were downloaded in the form of a Microsoft Excel spread sheet.

4.1.2. The training intervention (Appendix C)

The training intervention that was applied to test the extent to which a structured training intervention could enhance the understanding, perceptions, knowledge, and skills of LIS professionals was the RDM workshop hosted by NeDICC, which took place on the 11th of August 2015 at the CSIR.

The focus of the workshop was an introduction to and an overview of the value of RDM. Speakers/presenters included Joy Davidson and Sarah Jones from the DCC in the UK; Jim Mullins from Purdue University in the USA; and Elias Makonko and Anwar Vahed from SA. Most of the sessions were informational rather than practical.

The workshop was a day long (eight hours long), with breaks in between. At the start of the workshop Joy Davidson introduced the topic of RDM and explained the benefits and drivers of managing research data. Jim Mullins went on to discuss the impact that RDM has had on the profile of the library at Purdue University. The attendees were then given a small exercise called as the CARDIO quick quiz, to assess the RDM readiness of their institutions. After the exercise Elias Makonko described the state of RDM in SA and he was followed by Anwar Vahed from the Data Intensive Research Initiative of South Africa (DIRISA) who explained the DIRISA initiative and the impact it was expected to have on RDM in South Africa.

Joy Davidson shared a few of the tools and services offered via the DCC that organisations can reuse in their RDM endeavours and Sarah Jones demonstrated the data management tool DMPOnline. Towards the end of the workshop the attendees were asked to develop an RDM roadmap for their institutions before the ending discussion and closure.
The attendees had the opportunity to interact with the speakers and ask questions or make comments throughout the day.

Overall, the content covered during the RDM workshop was similar to the content of the in-house RDM training opportunities identified in section 2.7.2 of the literature review. The similarity of the workshop to the training opportunities is obvious because the keynote, presenters are affiliated with the DCC in the UK. The training materials used were based on the UK's DCC or JISC. Unique features of this particular workshop were the presentations about the state of RDM in SA and DIRISA and its implications for RDM in SA.

4.2. Themes identified for data analysis

Data was analysed and categorised into themes. The themes identified were the following:

- RDM policies at participants' institutions
- RDM services at participants' institutions
- Roles and responsibilities for offering RDM services
- Understanding of RDM
- Disciplinary background necessary for RDM
- Knowledge and skills gaps
- Feedback on the training intervention

Demographic information was collected on both questionnaires. The results of the demographic information are presented in section 4.3.1. Where possible, the percentages were rounded off to the whole number in all findings.

4.3. Data findings and analysis

The data collected using the online questionnaires are analysed below. The data is analysed according to the themes listed above.

4.3.1. Demographic information

The demographic information was asked in section two of the questionnaires (question 2.1-2.4). The participants were asked about their age, institution, position at institution and their years of experience.

a. Age and years of experience

The majority of the participants from the first observation (53%; 16/30) were 51 years and above, followed by 31-50 year olds (33%; 10/30) with the minority of the participants (13.3%; 4/30) being under 30 years of age. Fifty-two per cent (16/21) of the participants had over 21 years of experience. Almost 24% (23.3%) (7/21) of the participants had between 11-20 years of experience and, again, the minority of the participants (23.3%; 7/21) had less than 10 years of experience. As the researcher expected, participant's years of experience correlated with the age of the same participants.

The second questionnaire was answered by twenty-one (21) of the attendees. Close to 50% (48%) (10/21) of those who responded were 51 years and above, followed by 31-35 year olds (33%; 7/21) and only 19% (4/21) of the participants were under the age of 30. The participants' years of experience was evenly split across the three (3) categories – under 10 years of experience (33%; 7/21); 11-20 years of experience (33%; 7/21) and over 21 years of experience (33%; 7/21). The information set out here was captured in Table 4.2 below.

Table 4.2: Demographic information

	Observation 1 (N/31)	**Percentage**	**Observation 2 (N/21)**	**Percentage**
Age of participants				
Under 30	**4**	**13.3%**	**4**	**19%**
31-50	**10**	**33.3%**	**7**	**33%**
51 and above	**16**	**53.3%**	**10**	**48%**
Years of experience				
Under 10 years	**7**	**23.3%**	**7**	**33.3%**
11-20	**7**	**23.3%**	**7**	**33.3%**
Over 21	**16**	**53.3%**	**7**	**33.3%**

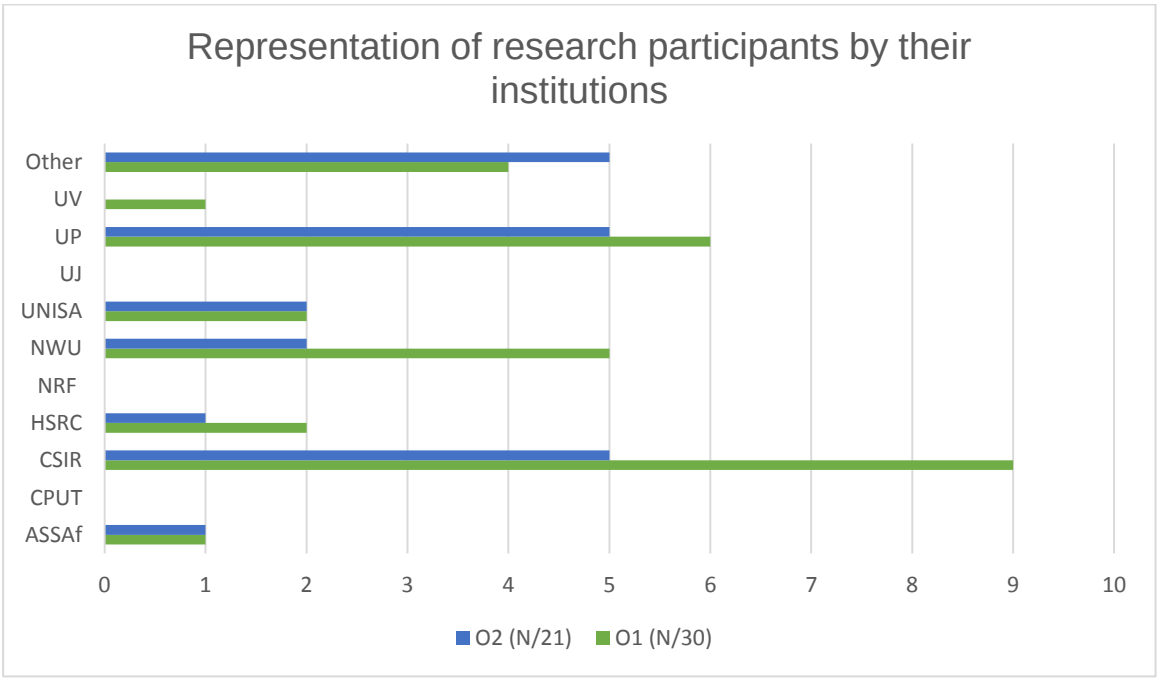
The table above indicates that the majority of the respondents were in the age category of 51 and above for both observations. From the first observation, the majority of the participant's experience fell in the over 21 category. Surprisingly, in the second observation, the participants' years of experience was evenly distributed across the three categories. The researcher assumes that participants aged 51 years and above could have started their careers late or they could have changed careers after working in a different profession.

b. Institutional representation

When answering the first questionnaire, the participants came from a varying institutions, with 30% (9/30) of participants from the CSIR-IS. Participants from UP made up 20% (6/30) of the participants and individuals from North-West University (NWU) made up 17% (5/21). It is unsurprising that the majority of the participants were from the CSIR as the workshop was hosted at the *CSIR's Knowledge Commons*. The response rate/attendance of the NWU

employees was surprising as the distance from the institution to the venue is far. Individuals from University of South Africa (UNISA) and the Human Sciences Research Council (HSRC) each made up 7% of the participants. The Academy of Science South Africa (ASSAf), and University of Venda (UV) each made up 3% of the participants. There was an "other" option for participants whose institutions were not listed. Sabinet (7%; 2/30); Department of Trade and Industry (DTI) (3%; 1/30) and Agricultural Research Council (ARC) (3%; 1/30) made up the rest of the group.

The individuals from the CSIR-IS and UP made up the highest number of participants for the research. Individuals from these two institutions each - made up 23% (5/21) of the participants followed by individuals from UNISA who made up 10% (2/21) of the research participants. Individuals from ASSAf, HSRC, NRF and NWU – each made up 5% (1/21) of the total participants. Other participants were from Stellenbosch University (5%; 1/21); Sabinet (10%; 2/21); DTI (5%; 1/21) and the ARC (5%; 1/21). The representation of the participants by their institutions is illustrated in Graph 4.1 below.



Graph 4.1: Representation of participants by their institutions

As stated previously (section 4.1.1.) the workshop was targeted at the LIS professionals in the Gauteng region, it was unsurprising that the majority of the participants were from Gauteng based campuses. It was positive to observe interest beyond the Gauteng region, e.g. LIS professionals from Stellenbosch University and North-West University.

c. Job positions

The job titles of the participants varied – even across both observations. These can be loosely categorised as management positions, LIS-related positions, data-related positions, and IT-related positions. In some cases, there were job titles that fitted in more than one category.

- **Data-related positions:** examples of the participants' data-related job positions were *data curator*, *manager, data management, data librarian, and research data curator.*
- **IT-related positions:** examples of the participants' IT-related job positions were: systems analyst, systems administrator, database operations manager and manager: information systems.
- **LIS-related positions:** most of the participants had LIS-related job titles ranging from indexer, cataloguer, records management and archival services, information specialist, information officer and institutional repository administrator, and information assistant.
- **Management positions:** examples of participant's management jobs were: manager: data management, knowledge manager, database operations management, and portfolio manager.

As expected, the majority of the participants were LIS professionals. It was surprising to see the number of professionals in data-related positions. Considering the fact there are already professionals in this area, these professionals could help with capacity building at national level. If the above-mentioned professionals can work together in initiating RDM in their institution, knowledge will be exchanged.

4.3.2. RDM policies

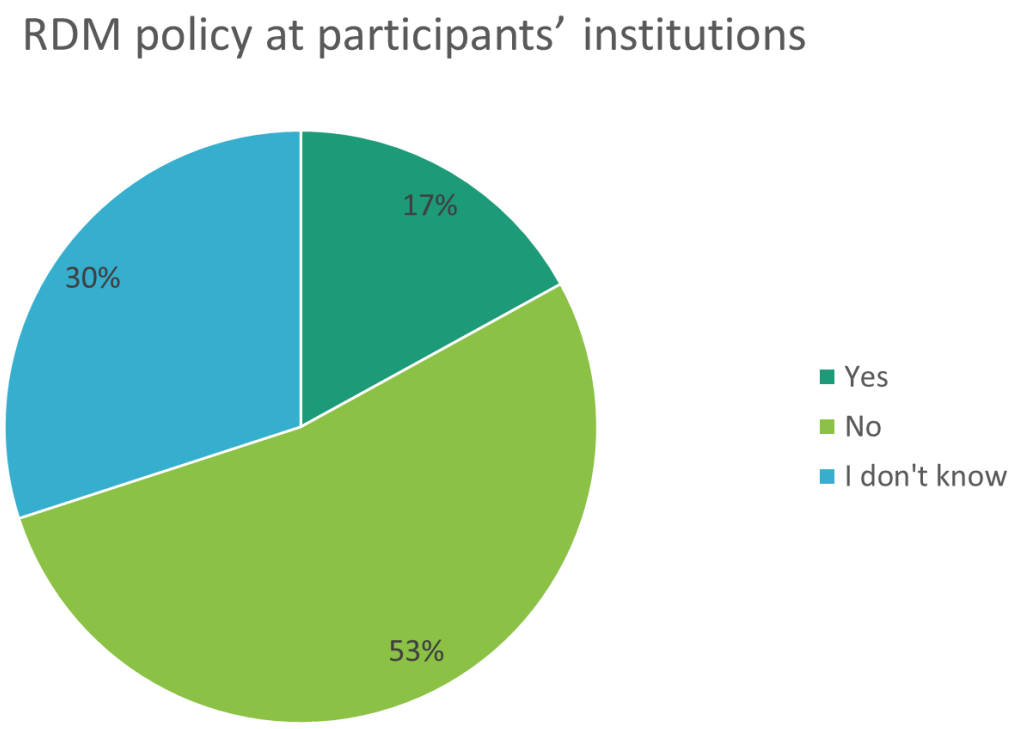
In section 3 of the questionnaire, the participants were asked about the RDM policies of their institutions and their opinions on what an RDM policy should address. Their answers are discussed below.

RDM policies at participating institutions

In question 3.1, the participants were asked whether their institutions had RDM policies. They could choose one of the following answers: "Yes", "No" and "I don't know".

Thirty percent (9/30) of the participants indicated that they did not know whether their institutions had such a policy while only 17% (5/30) said their institutions had such policies.

The majority of the participants, 53% (16/30) indicated that their institutions did not have RDM policies as depicted in graph 4.2. below.



Graph 4.2: RDM policies at participants' institutions

The questions about the RDM Policy (question 3.1 and 3.2) were only asked in the first questionnaire.

The fact that the majority of the participant's institutions did not already have RDM policies presents a challenge to the roll out of the initiative. Without a policy, the researchers at the institutions will not be compelled to manage their research data. Those that do want to manage their research data will most probably not have guidelines and standards to guide them. Without a policy, no institutional entity/entities will be held accountable for the task of ensuring data is managed properly.

Contents of the RDM policies

The participants were asked what aspects they thought an RDM policy should address. Five of the participants (16%) did not answer the question directly, leading the researcher to assume that they did not know enough about the topic to contribute. The suggestions the participants provided were divided into seven categories and summarised as follows:

RDM drivers: the policy should outline the objectives of RDM and describe why it is important. The benefits should also be mentioned to all stakeholders.

RDM responsibility: the policy should outline all the stakeholders and their roles and responsibilities, preferably aligned to the RDM lifecycle for clarity.

RDM process/procedure: the policy should discuss the entire RDM process/lifecycle, including the following areas: data-management planning, data capture, data curation/management, retention, archiving, sharing and publishing.

RDM guidelines, best practice standards: issues surrounding which data to store, quality control, ensuring accessing and increasing/improving visibility should be addressed in the policy. In addition, the policy should guide stakeholders on open access publishing, rights and ownership of data, security, and retention periods.

Governance: ethical considerations, legal compliance, ownership and rights of data creators and data re-users, and confidentiality issues should be addressed in the policy.

IT infrastructure: the policy should state where research data should be stored, how to ensure interoperability and what the software requirements for managing research data are.

Advocacy: creating awareness and providing training to the researchers.
The participants showed great understanding of what an RDM policy should entail. Although some wanted the policy to be as detailed as possible, others wanted the policy to have as little detail as possible. The categories identified above should be discussed in any RDM policy or RDM guidelines as part of a larger policy in an institution. It was seen as especially important to include the responsibility and the process/procedure to follow when managing research data

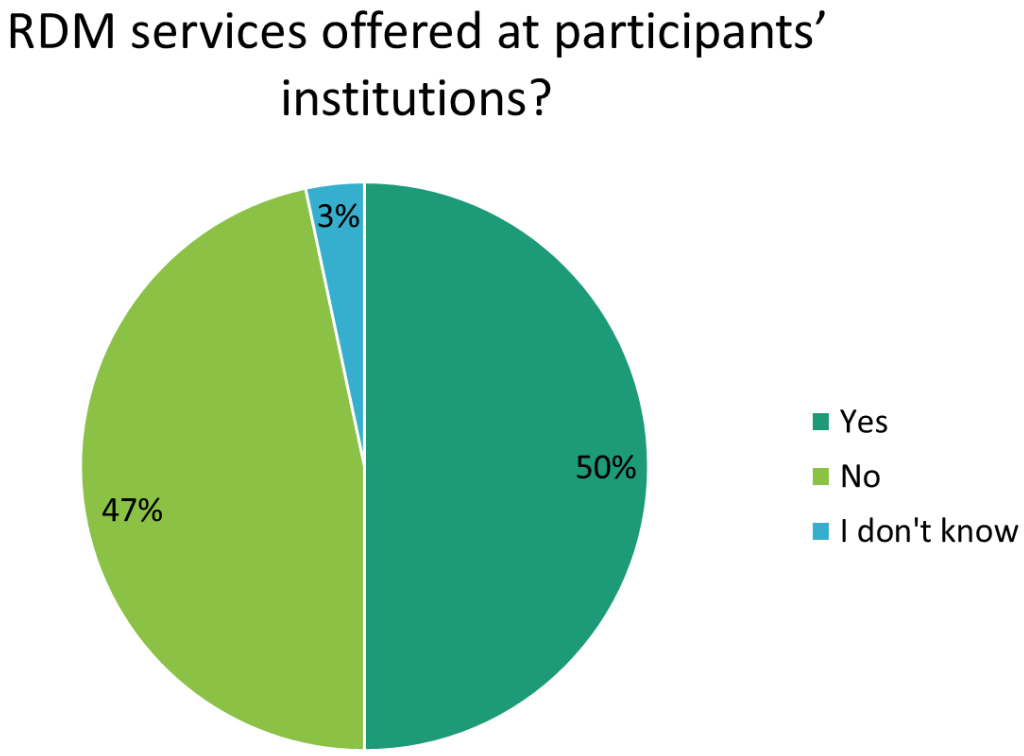
4.3.3. RDM services

The participants were asked whether their institutions offer RDM services and what they think RDM services should entail. The answers they gave are set out below.

RDM services offered by institution in attendance

In question 3.3, the participants were asked whether their institutions offered RDM services. The participants could answer: “Yes”, “No” or “I don’t know”.

Most of the participants (50%; 15/30) answered “Yes”, indicating that their institutions do offer RDM services. Forty-seven percent (14/30) of the participants answered “No”, while only 3% (1/30) of the participants indicated that they did not know. These findings are presented in the graph below.



Graph 4.3: RDM services offered by participant's institutions

It was interesting to observe that a higher percentage of the participants indicated that their institutions offered RDM services to their researchers because a large percentage of the participants indicated that they do not yet have policies. This seeming contradiction could mean that the participant’s institutions are offering RDM services without compelling the researchers to adhere to a policy.

RDM services that institutions should offer to researchers

The participants were asked to list RDM services they think their institutions should offer. The options they provided were categorised into three different types of services. Table 4.4. presents a summary of the participants’ answers.

Table 4.3: Categories of RDM Services

Infrastructural services	Consultation services	Applied services
Providing data storage space Providing IT hardware Providing cloud services Developing a repository (institutional or data)	Policy development Data management planning Offering advice on file formats Publishing Offering advice on database design Offering advice on data modelling Offering advice on data sharing Offering advice on data re-use rights Offering open access Offering advice on metadata standards	Providing technical support Data curation / actively managing data Providing preservation services Providing training Creating awareness Administering metadata to research data Digitisation services Designing data workflows and processes Training Creating guides/self-help manuals Facilitating RDM Creating awareness

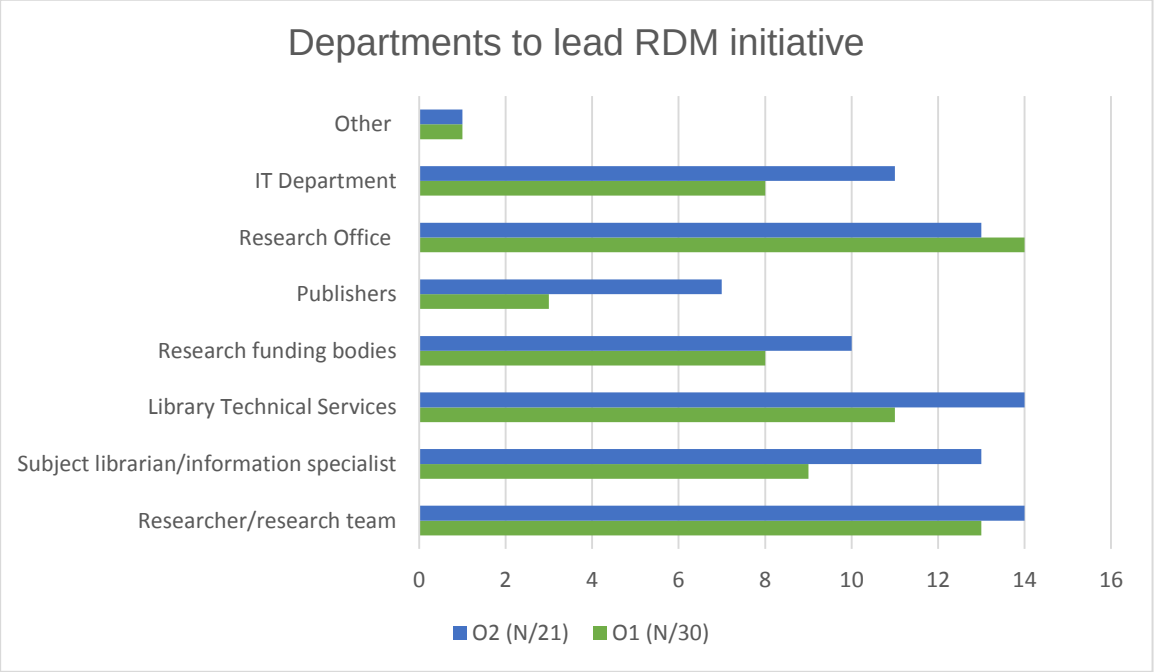
The ‘infrastructural services’ category represents the services that could ideally be provided by the institution, working together with the national bodies such as the NRF and the government. The infrastructure includes storage space, web-based tools, repository infrastructure, and other online resources. The ‘consultation services’ should include offering advice on various aspects of RDM as well as developing policies and guidelines for the nation, the institutions and their researchers. The ‘applied or technical services’ would be the actual *tangible* services that are offered to the researchers by their institutions. These services will require the involved professionals to interact with researchers to make them aware of RDM and what is expected of them while also providing RDM training to help researchers comply with policies and mandates.

4.3.4. Departments to lead the RDM initiative

Participants were asked which department they thought should lead the RDM initiative. They had the following options:

- Researcher or research team
- Subject librarians
- Library technical services
- Research office
- Research funding bodies
- IT departments
- Publishers
- Other

The participants could select more than one department to lead the RDM initiative. Graph 4.4 represents the participant's answers in both questionnaires.



Graph 4.4: Departments to lead RDM initiative

When answering the first questionnaire, 47% (14/30) of the participants chose the 'Research office' as the department to lead the RDM initiative. The 'Researcher or research team' was selected by 43% (13/20) of the participants and the 'Library technical services' by 37% (11/30). 'Subject librarians' was selected by 30% (9/30) of the participants and the 'Research funding bodies' and the 'IT department' were each selected by 27% (8/30) of the participants.

'Publishers' was only selected by 10% (3/30) of the participants to lead the RDM initiative. In the 'Other' option, a participant added that product managers should lead the RDM initiative.

When answering the second questionnaire the participants' thoughts on which department should lead the RDM initiative altered slightly. Sixty-seven percent (67%) (14/21) of the participants thought that the 'Researchers/research teams' and the 'Library technical services' should lead the RDM initiative. Sixty-two percent (62%) (11/21) of the participants selected 'Subject librarians' and the 'Research office' to lead the RDM initiative. Fifty-two percent (52%) (11/21) of the participants selected, the 'IT Department' and 48% (10/21) selected the research-funding bodies. Again, the publishers were the least preferred by the participants with only 33% (7/21) participants selecting these departments. The differences seem sharper than they actually are because the number of participants differs for the two observations. In the 'Other' option, one participant added that the 'Copyright office' should become involved in RDM initiatives.

From the discussion above it is clear that the participants thought that the 'Researchers/research teams' or the 'Research office' should lead the RDM initiative. Less participants thought that the 'Library technical services' or 'Subject librarians' should lead the RDM initiative. These finding differed from the literature (see section 2.5).

When the participants answered the second questionnaire (after the RDM workshop), their answers altered slightly to correlate with the literature. More participants thought the 'Library technical services' and 'Subject librarians' should lead the RDM initiative. The participant's preference for the 'Researcher or research team' to lead the RDM initiative also increased.

In this case, it was clear that the training intervention did impact on the participant's perceptions regarding the responsibility for RDM.

Services the selected departments should offer

The participants were asked what services they thought the departments they had selected should offer. This question was asked in both observations. The responses are provided in Table 4.

Table 4.4: Responsibilities of the leading departments

First questionnaire (Observation 1)	Second questionnaire (Observation 2)
Researcher or research team	
Help with discipline-specific best practice Plan data management Organise and manage their data Collect information [for research] Create datasets Use web-based tools to manage research Work with subject librarians	Create datasets Practice RDM as part of their research process Be accountable for managing own data Submit their data [to repositories]
Subject librarians/information specialists	
Advocate for RDM Provide RDM guidance, training and workshops Content management Assisting with DMPs Help develop RDM standards and procedures Overall management and curation of data Work with researchers throughout the research process	Advocate for RDM Manage data Provide data storage and retrieval services Be knowledgeable about the RDM process Help develop RDM policies Have subject knowledge Help researchers find existing data sets
Library technical services	
Provide and train researchers to use web-based tools Metadata administration Offer guidance to researcher Create guidelines to standardise the RDM process Overall management of preservation, archiving, classification Collect information from researchers Provide access to research data	Overall management of data – preservation, archiving and indexing. Provide access to data/datasets Offer advice on DMPs, data citation, finding data for re-use, locating data repositories Provide guidelines for managing research data.

Research office	
Participate in policy development Participate in developing discipline-specific standards and best practice Provide guidelines to institutions to ensure compliance and monitoring of institutional and/or funder policies Coordinate RDM service with an option to outsource to other departments Provide or ensure availability of storage services Negotiate with other stakeholders for changes to existing mandates Assist with DMPs Offer data management services such as archiving and preservation Collect information Provide and train researchers to use web-based tools Offer advice on: database design; data modelling; file formats; data sharing; and data re-use rights Offer a repository service	Participate in development of RDM policies and guidelines Assist researchers with DMPs Coordinate and manage RDM service Provide the necessary technology and infrastructure for RDM
IT Department	
Offer guidance to researchers Technical services such as organising, naming and structuring data. Data storage, security and backup Archiving and classification Providing web-based tools Developing IT infrastructure Leadership and coordination of RDM service	Develop IT infrastructure Data storage and backup

Data management Needs assessment for researchers	
Research Funding Bodies	
Providing RDM guidelines to institutions to ensure compliance Providing financial support Developing policy Providing training Discipline specific repository services	Financial contributions to aid in developing IT infrastructure and RDM skills Make past datasets available to researchers Lead in developing policy Provide incentives for compliance
Publishers	
Offer guidance to researcher	Make financial contributions to develop IT infrastructure and RDM skills
Other	
Product managers	Copyright office

Researcher or research teams

The participants thought that the researchers or research team should make RDM part of their research process and they should be accountable on their own data management. As much as it is the responsibility of the stakeholders to facilitate RDM in institutions the people that benefit the most are the researchers. These benefits stem from increased research impact/citations and authenticating research findings (see section 2.3). The participants also thought that the researchers should participate in their institution's RDM initiatives – such as developing policies and planning for RDM services as these initiatives will directly affects them.

Subject librarians/information specialists

Most of the participants thought that the information specialist/subject librarians should offer more of the consultation and applied services. These services include advocacy, overall management of data, offering advice and providing training to researchers. The subject knowledge of the subject librarians would allow them to work directly with researchers/research teams.

Library technical services

The participants also showed a strong preference for the Library Technical Services to lead the RDM initiative in both observations, suggesting that this department should offer the following services: overall data management, assisting with data management planning, accessing data, and providing support and training. The participants also thought that this department would be best suited for developing RDM processes and standards.

Research office

The participants thought that the Research office should coordinate and manage the RDM service in institutions. The Research office should also develop guidelines and best practice standards. The Research office should also ensure that there is sufficient infrastructure and monitor compliance. In summary the participants expect the Research office to liaise with the Research funding bodies and Researchers/research teams in their institutions to facilitate RDM in their institutions.

Publishers

It was interesting to note that the participants thought the publishers should make financial contributions to enable RDM. This was expected of the research funding bodies as they are the departments enforcing these policies and they receive funding from the government. Another service attributed to publishers was to offer guidance to the researcher. This is actually an acceptable idea (especially for data journals). Because publishers are already involved in offering guidance to researchers about publishing their research output it is likely that the researchers would comply with publishers.

Research funding bodies

Similar to the publishers, the participants thought that the research funding bodies should provide financial support to develop national RDM infrastructure, develop policies and guidelines and offer training to develop RDM workforce capacity. Another service that the participants would like the research funding bodies to provide is discipline-specific data repositories.

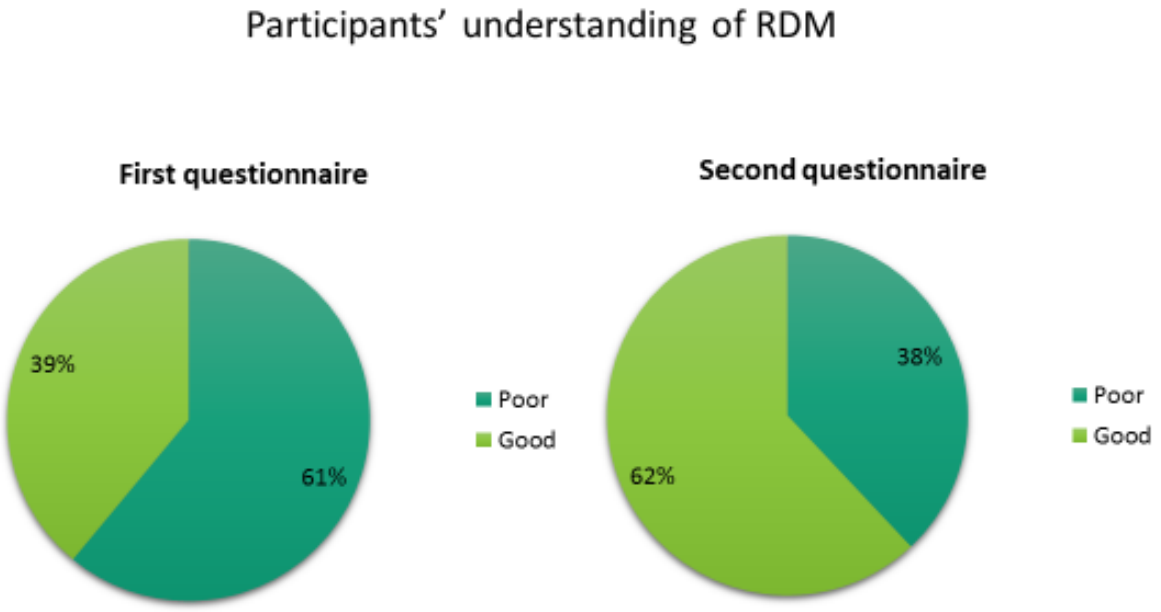
Other

The participants were provided with an option of other. When answering the first questionnaire, one participant indicated that he/she thought product managers should lead the RDM initiative. On the second questionnaire, a participant suggested that the copyright

4.3.5. Understanding of RDM

The participants were asked to rate their understanding of RDM on a scale of 1-4. One was considered poor and 4 was considered excellent. The participants who rated their understanding as 1 or 2 were categorised as having poor understanding of RDM. The participants who rated their understanding as 3 or 4 were categorised as having good understanding of RDM.

From the first questionnaire (before the workshop), the majority of the participants 61% (18/30) perceived their understanding of RDM as poor whereas 39% (12/30) of the participants perceived their understanding as good.

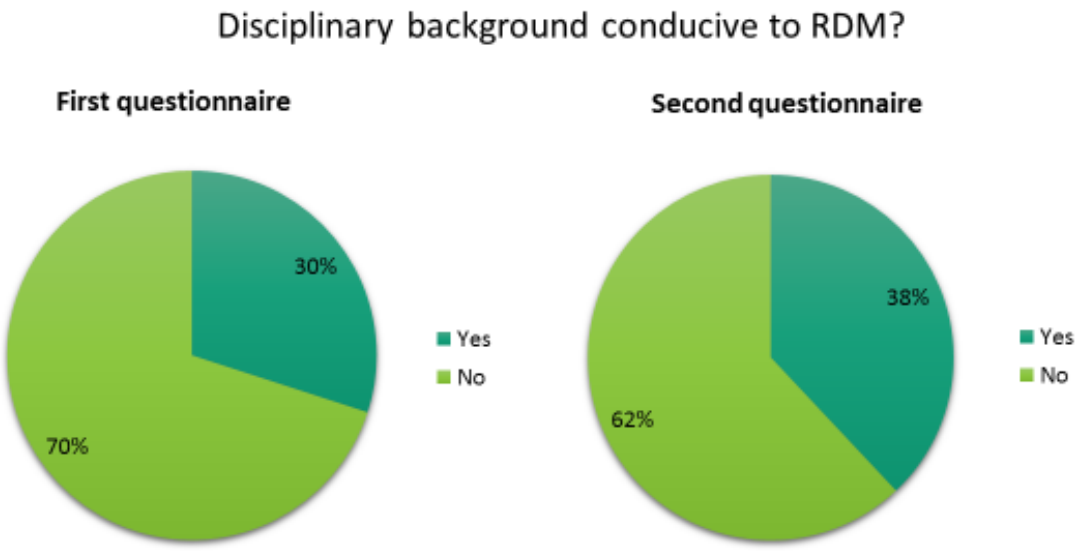


Graph 4.5: Understanding of RDM

When observed after the workshop had taken place, 62% (13/21) of the participants rated their knowledge as good and only 38% (8/21) thought they still had poor understanding of RDM. These results indicate that the workshop had a positive impact on the participants' RDM understanding/knowledge. Although there was a definite positive impact the workshop unfortunately was not sufficient to convince all participants that they had a good understanding of RDM.

4.3.6. Disciplinary background to offer RDM services

The participants were asked if they thought their disciplinary background was adequate for offering RDM services/advice. When answering the first questionnaire, 38% (11/30) of the participants replied ‘Yes’ and the rest of the participants (62%; 19/30) replied 'No’ . The question was asked again in the second questionnaire. There was a minor shift in the participant’s perception. Thirty-eight percent (38%) (8/21) replied “Yes”; 62% (13/21) replied “No” as depicted in Figure 4.6 below.



Graph 4.6: Disciplinary background conducive to RDM

When answering both questionnaires, the most cited reasons for answering “No” to this question was lack of formal RDM-related education and lack of practical experience in research and RDM. The participants mentioned that there is an overwhelming volume of information available on the topic of RDM. It is very easy to experience information overload. A participant mentioned that there is a need for summarised guidelines for institutions. Another participant indicated that further research needs to be conducted before the service could be offered.

When answering both questionnaires, the participants who replied “Yes” to the question indicated that their tertiary education makes them qualified to offer RDM services. Practical experience with managing data and that RDM was part of their current job responsibilities were also seen as reasons to be qualified to support RDM. Others said they had worked with related software such as content management systems (CMS) and repository software.

Interestingly, the issue of personal interest came up – one participant said he/she did not have the experience with RDM but that he/she did have personal interest in the subject.

4.3.7. Gaps in knowledge areas and skills sets of the participants

The participants were asked to rate their level of knowledge and skills in RDM-related areas.

4.3.7.1. Knowledge areas

The participants were asked to rate, on a scale of 1-4, the level of their knowledge in the following areas:

- Awareness of research data management
- Long-term preservation
- Policy and governance of data
- Copyright and other IP rights
- Metadata standards for data
- Trusted repositories
- Funder mandates
- Research process/lifecycle

This question was asked in both questionnaires. Table 4.5 represents a summary of their responses. To make better sense of the data, the participants' responses were categorised into two: poor and good. The participants who rated their knowledge a 1 or 2 were considered to have a poor understanding/knowledge of the RDM knowledge area and the participants who rated their knowledge as a 3 or 4 were considered to have good understanding/knowledge of the RDM knowledge areas. The total number of participants in both categories was then converted to percentages to compare the first and the second observation.

Table 4.5: Participant's perceived knowledge of RDM

RDM Knowledge Area	Poor knowledge		Good knowledge		Difference in perception
	Q1 (%)	Q2 (%)	Q1 (%)	Q2 (%)	
Awareness of research data management	50	33	50	67	17
Long-term preservation	63	52	37	48	11

Policy and governance of data	77	71	23	29	6
Copyright and other IP rights	60	48	40	52	12
Metadata standards for data	57	52	43	48	5
Trusted repositories	53	62	47	38	-9
Funder mandates	90	67	10	33	23
Research process/lifecycle	50	19	50	81	31

The first time the participants were observed, their perception of the RDM knowledge was mostly poor because the percentage of participants who rated their knowledge as 1 or 2 was 50% and above.

- Awareness of RDM:

Half of the participants (50%; 15/30) perceived their awareness of RDM as poor and the other half perceived their awareness of RDM to be good when answering the first questionnaire. After the intervention, 33% (7/21) of the participants indicated that they had poor awareness of RDM while 67% (14/21) of the participants perceived their knowledge of RDM to be good. The workshop, therefore, resulted in a 17% improvement in this area.

- Long-term preservation

At first, 63% (19/30) of the participants perceived their knowledge of long-term preservation to be poor while 37% (11/30) perceived their knowledge in this area to be good. After the intervention, there was a difference of 11%, with 52% (11/21) of the participants rating their knowledge as poor and 48% (10/21) rating their knowledge in this area. Again, a positive impact could be associated with the intervention.

- Policy and governance of data

A large percentage, 77% (23/30), of the participants perceived their knowledge of policy and governance of data to be poor while only 23% (7/30) had the opposite perception. After the intervention, a 6% difference was noted. Seventy-one percent (71%) (15/21) of the participants still perceived their knowledge in this area as poor and 29% (6/21) perceived their knowledge in this area to be good. The intervention can be seen to have had a positive impact on the participant's knowledge in this area.

- Copyright and other IP rights

Sixty percent (60%) (18/30) of the participants considered their knowledge of copyright to be poor when answering the first questionnaire and 40% (12/30) rated their knowledge as good. A difference of 12% was noted after the intervention with 48% (10/21) of the participants who considered their knowledge in this area as poor while 52% (11/21) rated their knowledge as good. There was a 12% improvement in the participants' perception of their knowledge of copyright and other IP rights after the workshop. .

- Metadata standards for data

Fifty-seven percent (57%) (17/30) of the participants thought that they had poor knowledge of metadata standards for data and 43% (13/30) thought their knowledge in this area was good when answering the first questionnaire. After the intervention, a difference of 5% was noted in the second questionnaire where 52% (11/21) of the participants rated their knowledge as poor and 48% (10/21) rated their knowledge as good. The intervention had a positive impact in this area.

- Trusted repositories

The participants' knowledge of trusted repositories was 53% (16/30) poor and 47% (14/30) good when answering the first questionnaire. After the intervention, the participants' knowledge had negatively shifted by 9%. Sixty-two percent (62%) (14/21) had poor knowledge in this area and 38% (8/21) of the participants had good knowledge in this area. This was the only area that the intervention appeared to have had a negative impact.

- Funder mandates

This knowledge of funder mandates had the largest percentage of participants with poor skills (27/30; 90%). Only 10% of the participants had good knowledge of this area. After the intervention, a significant improvement was observed with 67% (14/21) of the participants who perceived their knowledge as poor knowledge and 33% (7/21) had good knowledge in this area. The intervention had a very positive impact in this area with a 23% increase in the number of participants who rated their knowledge as good.

- Research process

The participants knowledge of the research process/lifecycle was evenly split (15/30; 50%) between poor and good knowledge when answering the first questionnaire. The difference observed after the intervention was the greatest observed in all knowledge areas, with a 31% positive difference. After the intervention, 81% (17/21) perceived their knowledge to be good

in this area, with only 19% (4/21) still having poor knowledge. In this area the workshop appears to have had a very positive impact.

4.3.7.2. Skills categories

The data from the skills set was analysed in the same as in the knowledge areas. The list of skills presented to the participants were the following:

- Indexing [adding metadata] skills
- Online citation/referencing for data
- Data Management Planning (DMP)
- Preparing data for deposit to repository
- Data licensing
- Data appraisal and selection
- Organisation and documenting data
- Data storage and security
- Create guides and training materials for researchers
- Working with data repositories
- Managing data repositories

The method used to analyse the data in section 4.3.7.1. was applied to analyse the participants'' perceptions of their RDM skills. Table 4.7 presents a summary of the participants' RDM skills.

Table 4.6: Participant's perceived skill sets related to RDM

Skills categories	Poor skills		Good skills		Difference in perception
	Q1 (%)	Q2 (%)	Q1 (%)	Q2 (%)	
Indexing [adding metadata] skills	53	62	47	38	-9
Online citation/referencing for data	43	48	57	52	-5
Data Management Planning (DMP)	63	62	37	38	1
Preparing data for deposit in a	70	71	30	29	-1

repository					
Data licensing	93	86	7	14	7
Data appraisal and selection	87	81	13	19	6
Organising and documenting data	63	52	37	48	11
Data storage and security	50	62	50	38	-12
Create guides and training material for researchers	57	62	43	38	-5
Working with data repositories	67	67	33	33	0
Managing data repositories	70	62	30	38	8

- Indexing [adding metadata] skills

The percentage of participants who perceived their indexing skills to be poor was 53% (16/30) the first time observed and 62% (13/21) after the intervention. The percentage of participants that thought that they had good indexing skills decreased by 9% after the intervention

- Online citation/referencing of data

The percentage of participants who perceived their online data citation skills as poor when answering the first questionnaire was 43% (13/30). This meant that 57% (17/30) of the participants perceived their knowledge in this area as good. The percentage of participants with poor skills in this area increased to 48% (10/21) after the intervention. The percentage of participants with good online data citation skills was 57% (17/30) before; which decreased by 5% to 52% (11/21) after the intervention.

- Data Management Planning (DMP)

Sixty-three (63%) (19/30) of the participants perceived their DMP skills to be poor when first observed and 62% (13/21) after the intervention. The participants with good DMP skills increased by 1%.

- Preparing data for deposit

Seventy percent (70%) (21/30) of participants perceived their skills to be poor in this skills category. After the intervention, a 1% negative difference was observed. The percentage of participants with poor skill increased to 71% (15/21).

- Data licensing

Ninety-three percent (93%) (28/30) of the participants had poor data licensing skills before the intervention and 86% (18/21) after the intervention. The intervention changed the participant's perceptions of their data licencing skills positively by 7%.

- Data appraisal and selection

On the first questionnaire, 87% (26/30) of the participants perceived their data appraisal and selection skills to be poor. After the intervention, 81% (17/21) of the participants thought that their data appraisal and selection skills to be poor. The participants perceptions changed 6% positively. The intervention was effective in changing the participants' perceptions of their data appraisal and selection skills.

- Organisation and documenting data

On the first questionnaire, 63% (19/30) of the participants indicated that their skills in organising and documenting data were poor. Thirteen percent (13%) (11/30) of the participants thought, they had good data organisation and documentation skills. After the intervention, 52% (11/21) of the respondents had poor data organisation and documentation skills while 48% (10/21) thought the opposite. The intervention was effective in changing the participants' perception of their skills in their data organisation and documentation skills.

- Data storage and security

The participant's data storage and security skills was evenly split at 50% (15/30) poor and 50% good (15/30) before the intervention. After the intervention, 62% (13/21) of the participants perceived their skills in this area to be poor and 38% (8/21) had good skills in this area. The difference was 12% negatively.

- Creating guides and training

Fifty-seven percent (57%) (17/30) of the participants rated their skills in this area as poor, while 43% (13/30) rated their skills in this area as good on the first questionnaire. After the

intervention, the participants' perception of their skills in this area had shifted negatively – 62% (13/21) poor and 38% (8/21) good. This shift represented a 5% negative difference.

- Working with repositories

When answering the first questionnaire 67% (20/30) of the participants perceived their skills in this area as poor and 33% perceived their skills as good. After the intervention, no difference was noted.

- Managing repositories

70% (21/30) of the participants thought their skills in this area were poor while 30% (9/30) thought they had good skills in this area before the intervention. After the intervention, a decrease of 8% in the participants who thought they had poor skills in this area was observed, with 62% (13/21) of the participants perceiving their knowledge as poor and 38% (8/21) as good.

The workshop included very few skills-development opportunities. It therefore appears that the workshop may have caused a reality check to have taken place.

4.3.1. Observation 1 versus observation 2

First questionnaire/Observation 1 (Appendix A)

As was stated before: The first questionnaire was administered before the RDM workshop (intervention). The individuals that had registered for the workshop were asked to complete the questionnaire.

The questionnaire contained 16 questions in total. From that observation, the researcher found out that the majority of the institutions represented did not yet have policies in place to regulate RDM although some had already started offering RDM services to their researchers.

The participant's perceptions of their RDM understanding, knowledge and skills seemed quite poor at this stage. The participants were also of the opinion that they did not have the necessary disciplinary background to offer RDM services.

The types of services that the participants thought their institutions should offer included infrastructural services, consultation services and technical (applied) services. The participants were also of the opinion that the researchers and the Research office should lead the RDM initiative.

Second questionnaire/Observation 2 (Appendix B)

The second observation also asked 16 questions, five of which were about the workshop as such. The other 11 questions asked in the questionnaire were the same as those asked in the first questionnaire.

The responses to the second questionnaire revealed a definite change in the participants' opinions. The participants' opinions changed regarding who they thought should lead RDM. More of them thought that the subject librarians/information specialists should lead the initiative. The participants' perceptions of their understanding and knowledge of RDM changed in seven of the eight knowledge areas. The participants' perceptions of their RDM skills improved in five of the eleven skill categories.

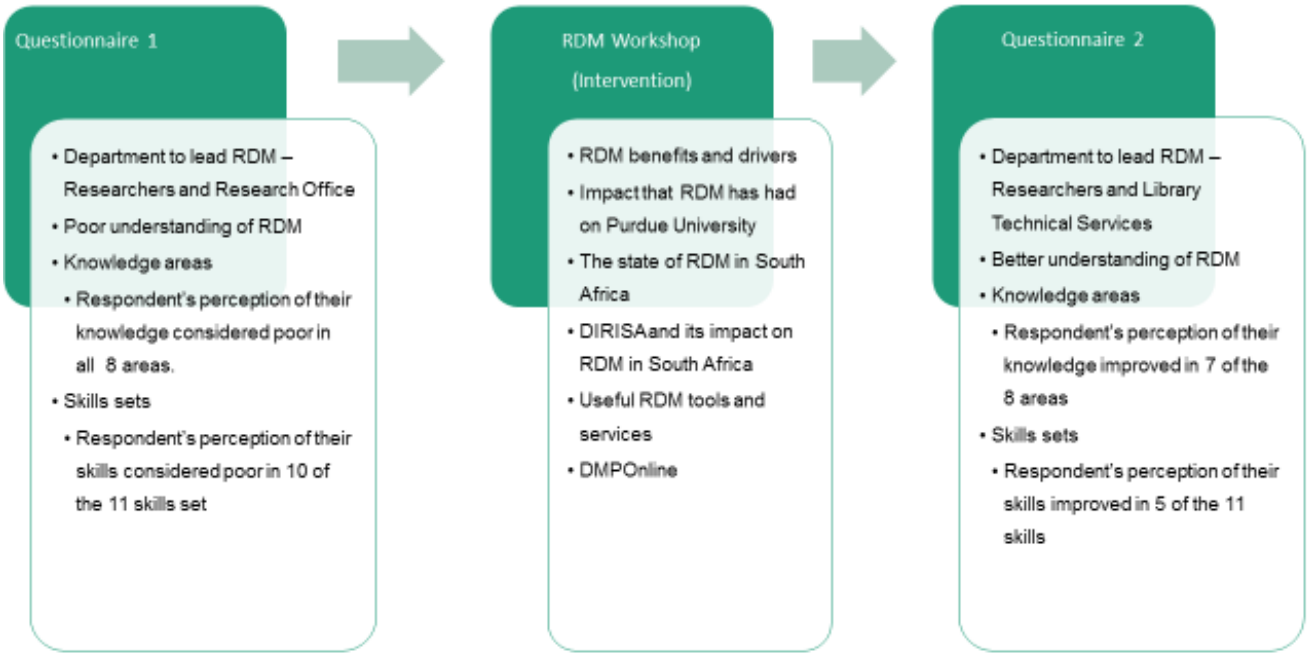


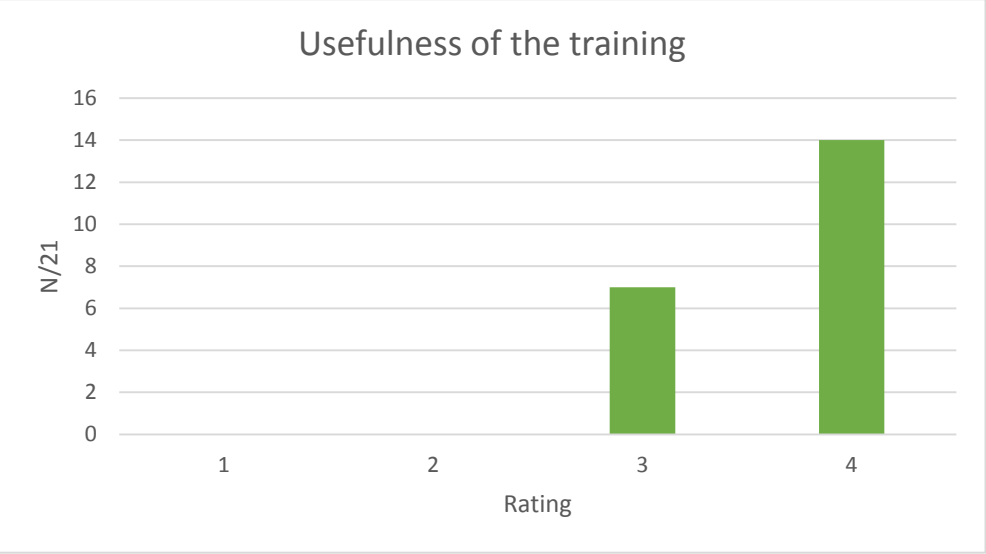
Figure 4.1: Changes resulting from the intervention

Figure 4.1 shows the most important changes resulting from the research intervention (RDM workshop by NeDICC). Overall it is possible to report that the workshop had a very positive outcome. The results of the questions that relate to feedback regarding the RDM workshop are presented in section 4.3.9.

4.3.2. Feedback from RDM workshop

The second observation asked questions specifically about the workshop. The responses are presented below:

The participants were asked to rate the useful of the training on a scale of 1-4. Graph 4.8 depicts the participants' response to the question about the usefulness of the workshop.



Graph 4.7: Usefulness of the training intervention

None of the participants gave a rating below three (3). Thirty-three percent (33% or 7/21) of the participants rated the training as three (3) and 67% (14/21) of the participants rated the training as four (4), leading the researcher to conclude that the workshop was seen as very to the participants.

The question that followed asked what the participants would have liked to be included in the workshop that wasn't addressed. Most of the participants (52%; 11/21) were satisfied with the workshop, some adding that the workshop gave them a greater understanding of the topic.

Forty-eight percent (48%) (10/21) of the participants added things they would have liked to have been included in the workshop. These were:

- More practical examples of RDM
- [More coverage on] data citation
- More examples from South Africa showcasing bad and good RDM practices

- Demonstration of data management
- Hand-outs of the presentation (the speakers from the DCC had an accent and it was hard for some to follow)
- More information about repositories
- More detail regarding data curation
- More time on indexing

One participant thought the content was too much for the duration of the workshop. The participants suggested extending the workshop over two days to give attendees time to grasp the concepts and apply their newly acquired knowledge.

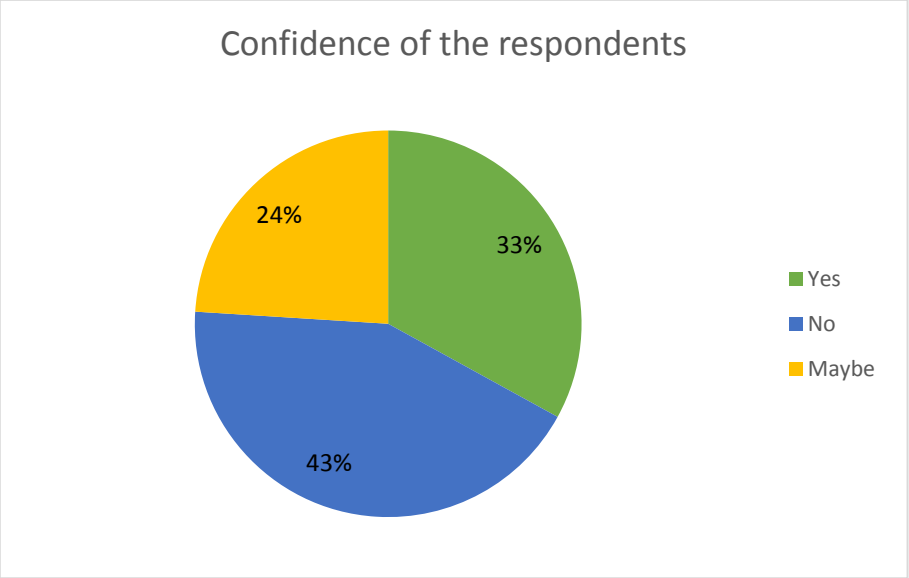
The participants were also asked to mention what they found useful about the workshop. The participants indicated that they learnt more about RDM, the benefits, the drivers, the role players and the skills needed. The participants found it useful to learn about the role that the library is playing/will play in RDM. One participant showed excitement for RDM and remarked:

"*This is definitely the future for librarians and I'd like to be a part of that future.*"

Another participant agreed that RDM will become a part of the librarian's world in the future but added that the role doesn't necessarily belong there.
Learning about RDM at a national and international level was also very useful. Although not officially part of the programme, the work done by the CPUT library to facilitate RDM at their institution was discussed briefly. The participants also found this useful, as was the presentation about DIRISA. The participants said that it helped to see was what was being done by South Africans HEIs and research councils.
Some participants also mentioned that they had learnt a lot about the tools and resources already available online. Specific mention was made of DCC training material. The exercises were found to be useful by the participants. An exercise towards the end of the workshop where the attendees were asked to complete an exercise to develop an RDM roadmap for SA institutions, was seen as very useful. The participants indicated that they had learnt the importance of planning at an institutional level.

The participants were then asked if they felt confident that they could assist with RDM-related questions and queries from research staff. This was an open-ended question. The answers were categorised. 33% (7/21) were confident that they would be able to assist researchers. 24% (5/21) were somewhat confident and the majority 43% (9/21) did not feel

confident that they could answer RDM questions from research staff. Graph 4.8 reflects the responses.



Graph 4.8: Confidence of the participants

Other comments that were captured were:

- The participants would like notification of future RDM workshops
- Future workshops – more detailed and broken up to focus on each aspect of the RDM process
- Link to the University of Pretoria's presentation by Jim Mullins was not very clear – time could have been used to hear more from the DCC speakers
- Need more SA examples of RDM being implemented
- Examples of archived data on repositories
- Video illustrating entire RDM process to help beginners understand

The feedback received from the participants was positive. The participants rated the usefulness of the training as very useful and a difference was observed in the quality of the answers provided by the participants in the second observation. In addition, a higher percentage of the participants indicated that their understanding of RDM had increased after the training intervention. This improvement further demonstrates that the training intervention had an effect on the participant's perceptions of their RDM understanding, knowledge and skills.

The additional comments that the participants provided were also quite useful and could be used when planning future training interventions. One of the most common comments was

that the information provided was too much for a day-long workshop. Having shorter, more focused training interventions would allow for more professionals to attend as it might be cheaper and institutions would be able to allow more of their staff members to attend.

It is worrying that many of the participants (43%) still do not have the confidence to offer RDM services. This is an indication that there is a need for more training interventions, perhaps focusing on developing the necessary skills and building the LIS professional's confidence.

4.4. In summary

The findings of this research illustrated the current situation in the participant's institutions. It was found that the majority of institutions represented did not have policies in place even though they had started offering some RDM services. The participants also showed a strong preference for the researchers to partake in RDM initiatives, working together with the library technical services, the subject librarians/information specialists and the research office.

It was also found that the training intervention applied in this research positively enhanced participants' perceptions of their RDM understanding and knowledge. The participants' perceptions of their RDM skills were mixed after the intervention.

CHAPTER 5: RECOMMENDATIONS AND CONCLUSIONS

5.1. Introduction

The previous chapter presented the findings from the data collected during the research. This chapter discusses the findings in relation to the central research question and research sub-questions discussed in Chapter 1 and Chapter 3. In this chapter, recommendations are made as to what can be done to develop training events to enhance the RDM understanding, knowledge and skills of LIS professionals. Recommendations are also made for future workshops and further studies on RDM skills development in South Africa.

5.2. Research findings

The research investigated the extent to which a structured training intervention could enhance the RDM understanding, knowledge and skills of LIS professionals at South African HEIs. Five sub-questions were formulated to help complete the investigation. These sub-questions were answered using the literature reported in Chapter 2 and the data collected using the embedded design discussed in Chapter 3 to collect the data presented in Chapter 4.

Section 5.2.1 answers the first research sub-question.

5.2.1. What is seen internationally as RDM?

Briefly described, RDM comprises of activities undertaken to keep data secure, accessible and re-usable. These activities can differ depending on the lifecycle model used. The most common activities performed in RDM are: planning, collecting, processing, analysing, managing, preserving, sharing, and re-using data. Both the RDM models and definitions are provided in Chapter 2 (see section 2.2).

5.2.2. Should libraries and LIS professionals play a role in their HEI's RDM efforts? What could this role be?

Should libraries and LIS professionals play a role in their HEIs RDM efforts?

The literature reviewed indicated that libraries and LIS professionals definitely have a role to play in their institution's RDM efforts (see section 2.5). At first, the results of the research contradicted the literature. When answering the first questionnaire, the participants preferred

other departments to lead the RDM initiative, particularly the 'Researcher of research team'. It was only after the participants attended the RDM workshop (training intervention) that they envisioned a role for libraries and LIS professionals in their institution's RDM efforts (refer to section 4.3.4).

The role of the library and LIS professionals

The participants thought that the 'Library technical services' could play a role in developing an institutional policy and providing RDM services. These should include:

- Actively managing data by offering services such as metadata administration, data preservation and archival;
- Assisting researchers with DMPs;
- Providing access to past data;
- Consulting with researchers and research teams; and
- Providing training to researchers.

The RDM workshop changed the participants' opinions on the role of libraries and LIS professionals. This change in opinion can be attributed, in particular, to the Skype presentation by Professor Jim Mullins on the impact that RDM has had on Purdue University Libraries. The researcher assumes that hearing about the successes that Purdue University Libraries had in their RDM efforts could have made the participants realise a potential role for themselves.

5.2.3. What RDM understanding, knowledge and skills do LIS professionals need for playing these roles?

The research found that most of the participants did not think they had the disciplinary background to initiate RDM in their institutions, citing lack of formal education and lack of practical work experience as the reasons (refer to section 4.3.6). In the literature, however, LIS professionals are considered the most competent professionals to play a role in their institution's RDM efforts (refer to section: 2.5.2).

Most authors (Heidorn, 2011; CILIP, 2014; Corrall *et al.,* 2013 and Cox & Pinfield, 2014) are of the opinion that LIS professionals are suited to take on this role due to their experience in managing information collections. It is also believed that the roles that the LIS professionals will potentially play in RDM efforts align with their current practices (Gold, 2008: online; Cox *et al.* 2012: online and Tiwari & Chand, 2014: 243).

From the literature (Antell *et al,.* 2014; Corral *et al.,* 2013; Charbonneau, 2013; Daland, 2015; Lyon, 2007; Brown *et al.,* 2014) the researcher was able to identify some of the RDM-related understanding, knowledge and skills that LIS professionals must have to enable and/or support RDM in their institutions. The participants were provided with these lists in both questionnaires. The participants were asked to rate their understanding, knowledge and skills. These findings were presented in section 4.3.7. Table 5.1 below, presents a summary of these findings as well as recommendations to overcome gaps in LIS professional's RDM skills and knowledge

Table 5.1: Understanding, knowledge and skills necessary for the LIS professionals

RDM requirement	Findings	Recommendation
RDM Understanding	The participants did not think that they had good understanding of RDM when they were answering the first questionnaire. The intervention (RDM workshop) did effect a change in this regard because more participants indicated that they had better/improved understanding of RDM when answering the second questionnaire.	This aspect is easy to address via a workshop and should be included in all training events.
Knowledge areas:		
Awareness of RDM Research process Long-term preservation Metadata standards for data Policy and governance of data Copyright and other IP rights Funder's mandate Research process/lifecycle	More than 50% of the participants rated their knowledge as poor in all areas when answering the first questionnaire. The intervention was successful in enhancing the participant's knowledge in seven (7) of the eight (8) areas.	Future interventions should focus on a specific aspect to reduce the possibility of information overload. Individuals can make use of the online training material, made available by the DCC, to learn more about RDM

Skills sets		
Indexing [creating and adding metadata] Data referencing and citation Organising and documenting data Creating guides and training Data storage and security Data licensing Data management planning Working with repositories Managing repositories	Even though no skills development training was included in the workshop participants appeared to gain a more realistic understanding of the skills set necessary for supporting and enabling RDM. The participants did not think they had the disciplinary background to offer RDM services. One of the reasons was that they did not have practical experience in working with data.	Training material to develop very specific skills needs to be developed for future interventions. Hands-on (practical) training is likely to be more effective in developing the RDM skills of LIS professionals.

5.2.4. What is the best way to obtain or improve the LIS professional's RDM understanding, knowledge and skills?

There are several strategies available to help LIS professionals improve their RDM understanding, perceptions, knowledge and skills. The LIS professionals can be offered formal and structured training or online, self-administered training. The formal training interventions could range from in-house training to graduate-level educational programmes. Some of the training interventions available were listed and discussed in Chapter 2 (see section: 2.7).

There is no one training intervention that can help LIS professionals acquire all the understanding, knowledge and skills they need. LIS professionals can try to obtain or improve RDM understanding, knowledge and skills by enrolling in a variety of workshops/training interventions that focus on different aspects of RDM. they may also choose to enrol in formal university-level degrees .

In SA, the NRF should work collaboratively with HEIs and publishers to develop appropriate training interventions to skills the LIS workforce. In addition these departments should try to

financially support interested LIS professionals who may enrol in formal educational programmes to improve their RDM knowledge and skills.

5.2.5. To what extent would a structured RDM training intervention enhance the RDM understanding, knowledge and skills of the LIS professionals?

Understanding

The training intervention enhanced the LIS professional's understanding of RDM (refer to section 4.3.5).

Knowledge

The training intervention was effective in enhancing the LIS professional's perception of their RDM knowledge. The intervention improved the knowledge the LIS professionals in seven (7) of the eight (8) knowledge areas (refer to section 4.3.7).

Skills

The training intervention was not as effective in enhancing the RDM skills of the LIS professionals. As stated in Chapter 4 (section 4.3.7) the training intervention had very few opportunities for skills development. The intervention did manage to provide the participants with an opportunity to reflect on their skills compared to the skills necessary for enabling and supporting RDM.

5.3. Research limitations

The research had the following limitations;

- The results of the research cannot be generalised as this was a singular, intrinsic case study; and
- There was a definite bias towards LIS professionals who live in the Gauteng Province.

5.4. Recommendations

5.4.1. Recommendations for improving the understanding and knowledge and skills of the LIS workforce

As discussed in Chapter 2 (section 2.7.2) there are various training interventions available to LIS professionals. It is up to individuals to identify their own knowledge and skills gaps before selecting an appropriate intervention. The DCC in the UK has made training material available to the public. Such availability makes it possible for interested LIS professionals to learn on their own. LIS professionals who qualify for university entrance may choose to enrol in an academic programme with a specific focus on RDM or related subject. Institutionally, the library (line managers) may choose to offer their employees the opportunity to attend specific training in preparation for their new role. Libraries can also develop in-house RDM training interventions using the freely accessible training material.

5.4.2. Recommendations for future workshops

Future workshops should be less general but more focused on one theme/topic per session to give the participants enough time to wholly understand what is being discussed. Attendees can be provided with hand-outs of the presentations beforehand so that they can follow the speakers (especially if the participants are not too familiar with the topic or the speakers speak with a different accent).

If possible, more SA examples should be shared during these workshops. It is encouraging for the attendees to hear what is being done within the borders of their country. For wider reach, it would be advisable to have the workshops virtually or electronically. The virtual sessions would be more suitable for enhancing the participants' RDM knowledge. For skills development, hands-on training should be developed, each session focusing on a particular skills category.

5.4.3. Recommendations for further studies

The researcher recommends that in future similar studies be conducted with a few modifications. The modifications could be:

- A similar study with a larger researcher population;
- A similar study comparing two or more training interventions; or
- A study investigating the soft skills needed to facilitate RDM.

5.5. Conclusion

The recently issued statement by the NRF has intensified the need for establishing RDM in HEIs. As discussed in previous chapters, libraries and LIS professionals are expected to play a role in their HEIs' RDM efforts. To play this role, the LIS professionals would need to enhance their current knowledge and skills. This research investigated if a specific RDM training event would enhance the participants' perceptions of their RDM knowledge and skills. The research found that the selected training intervention did enhance the participants' perceptions of some of the skills and knowledge. One could therefore then deduce that similar events would be useful – especially when the LIS professionals start on the RDM journey. The research did have some limitations, the most important of which was that the results of this research cannot be generalised. However, the results of this research could be used as a benchmark for further studies.

REFERENCES

1. Akers, K.G. and Doty, J. 2012. Differences among Faculty Ranks in Views on Research Data Management. *IASSIST Quarterly,* 36(2): 16-20. [Online] Available at: http://iassist.lishost.com/sites/default/files/iq/iqvol36_2_doty.pdf [Accessed on: 11 December 2014].
2. Antell, K., Bales, F., Turner, J. and Shults, B. 2014. Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries,* 75(4): 557-574. [Online] Available at: http://crl.acrl.org/content/early/2013/04/05/crl13-464.short [Accessed on: 11 December 2014].
3. Ashley, K. 2012. Research data and libraries: who does what? *Insights: the UKSG Journal,* 25(2): 155-157. Available at: http://uksg.metapress.com/index/14015N68L0515541.pdf [Accessed on: 11 December 2014].
4. Association of Research Libraries (ARL). 2015. *E-Research*. [Online] Available at: http://www.arl.org/focus-areas/e-research#.Vnkn-rZ96wU [Accessed on: 11 November 2015].
5. 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*. RLUK*. [Online] Available at: http://0-hdl.voced.edu.au.innopac.up.ac.za/10707/204093 [Accessed on: 11 July 2015].
6. Ball, J. 2013. Research data management for libraries: getting started. *Insights: the UKSG Journal,* 26(3): 256-260. [Online] Available at: http://uksg.metapress.com/index/VX30V77804671756.pdf [Accessed on: 11 December 2014].
7. Bradbury, S. and Borchert, M. 2010. Survey of eResearch practices and skills at QUT, Australia. *International Association of Scientific and Technological University Libraries, 31st Annual Conference. Paper 3.* Purdue e-Pubs, 21 June 2010.
8. Carlson, J.R. 2012. Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data. *Reference Services Review,* 40(1): 7-23. [Online] Available at: http://dx.doi.org/10.1108/00907321211203603 [Accessed on: 11 December 2014].
9. Carlson, J.R. 2014. The use of life cycle models in developing and supporting data services. In: Ray, J.M., Charlson (ed.) *Research Data Management: Practical*

strategies for information professionals. West Lafayette, Indiana: Charlson Insights in Library, Archival and Information Science.

10. Carlson, J.R., Johnston, L., Westra, B., and Nichols, M. 2013. Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project. *The International Journal of Digital Curation,* 2(2): 204-217. [Online] Available at: http://dx.doi:10.2218/ijdc.v8i1.254 [Accessed on: 29 January 2015].
11. Chad, K. and Enright, S. 2014. The research cycle and research data management (RDM): innovating approaches at the University of Westminster. *Insights: the UKSG Journal,* 27(2): 147-153. [Online] Available at: http://dx.doi.org/10.1629/2048-7754.152 [Accessed on: 12 December 2014].
12. Charbonneau, D.H. 2013. Strategies for data management engagement. *Medical reference services quarterly,* 32(3): 365-374. [Online] Available at: http://dx.doi.org/10.1080/02763869.2013.807089 [Accessed on: 12 December 2014].
13. Chartered Institute of Library and Information Professionals (CILIP). 2013. *Research Data Management Briefing Paper*. [Online] Available at: http://www.cilip.org.uk/sites/default/files/documents/Research%20data%20management%20briefing%20July%202014_0.pdf [Accessed on: 13 October 2014].
14. Clements, A. 2013. Research information meets research data management ... in the library? *Insights: the UKSG Journal,* 26(3): 298-304. [Online] Available at: http://uksg.metapress.com/index/YG3MQ143127037GW.pdf [Accessed on: 11 December 2014].
15. Coates, H.L. 2014. Building Research Data Services from the Ground Up: Strategies and Resources. *Journal of eScience Librarianship,* 3(1): 52-59. [Online] Available at: http://dx.doi.org/10.7191/jeslib.2014.1063 [Accessed on: 15 January 2015].
16. Corrall, S. 2012, Roles and responsibilities: libraries, librarians and data. In: Pryor, G. (ed.) *Managing Research Data*. London, UK: Facet Publishing.
17. Corrall, S., Kennan, M. and Salo, D. 2013. Research know-how for research support services: Preparing information specialists for emerging roles. *Proceedings of the American Society for Information Science and Technology,* 50(1): 1-4.
18. Corti, L., Van den Eynden, V., Bishop, L. and Wollard, M. 2014. *Managing research data: A guide to good practice.* Thousand Oaks, CA: Sage Publishers.
19. Cox, A., Verbaan, E. and Sen, B. 2012. Upskilling liaison librarians for research data management. *Ariadne*. [Online] Available at: https://0-www.ariadne.ac.uk.innopac.up.ac.za/issue70/cox-et-al [Accessed on: 11 November 2015].

20. Cox, A.M. and Pinfield, S. 2014. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science,* 46(4): 299-316. [Online] Available at: http://lis.sagepub.com/content/46/4/299.short. [Accessed on: 15 December 2014].
21. Creamer, A., Morales, M.E., Crespo, J., Kafel, D. and Martin, E.R. 2012. An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England. *Journal of eScience Librarianship,* 1(1): 18-26. [Online] Available at: http://dx.doi.org/10.7191/jeslib.2012.1006. [Accessed on: 29 January 2015].
22. Creswell, J.W. 2013. *Qualitative inquiry & research design: choosing among five approaches.* 3rd Ed. Los Angeles, CA: Sage Publishers.
23. Daland, H. 2015. Library Instruction – Not Just for our Users: Skills Upgrading for Librarians as a Way of Increasing Self-Confidence. *Liber Quarterly: The Journal of the Association of European Research Libraries,* 25(1): 2-17. [Online] Available at: http://liber.library.uu.nl/index.php/lq/article/download/10002/10598 [Accessed on: 22 August 2015].
24. Davenport, T. and Patil, D.J. 2012. Data scientist: Data Scientist: The Sexiest Job of the 21st Century. *Harvard business review,* 90: 70-76. [Online] Available at: http://www.tias.edu/docs/default-source/Kennisartikelen/harvard_data-scientist-the-sexiest-job-of-the-21st-century_2012.pdf?sfvrsn=0 [Accessed on: 15 August 2015].
25. Davidson, J., Jones, S., Molloy, L. and Kejser, U.B. 2014. Emerging Good Practice in Managing Research Data and Research Information within UK Universities. *Procedia Computer Science,* 33: 215-222. [Online] Available at: http://forskningsbasen.deff.dk/Share.external?sp=Sb1d5f839-3331-4ec9-9741-2ec6fd3e4bee&sp=Ska [Accessed on: 18 December 2014].
26. Dawson, C. 2009. *Introduction to research methods: a practical guide for anyone undertaking a research project.* 4th Ed. London, UK: Oxford.
27. 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.
28. Dietrich, D. 2010. Metadata Management in a Data Staging Repository. *Journal of Library Metadata,* 1(2-3): 79-98. Available at: http://www.tandfonline.com/doi/abs/10.1080/19386389.2010.506376. [Accessed on: 11 December 2014].

29. Donmoyer, R. 2008. Quantitative Research. In: Given, L.M. (ed.) *The Sage Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: Sage Publishers.
30. Donnelly, M. 2012. Data Management plans and planning. In: Pryor, G. (ed.) *Managing Research Data.* London, UK: Facet Publishing.
31. Gold, A.K. 2007. Cyberinfrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *Office of the Dean (Library).* [Online] Available at: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=lib_dean [Accessed on: 11 September 2015].
32. Gore, S.A. 2013. An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England. *Journal of eScience Librarianship,* 2(1): 20-24. Available at: http://dx.doi.org/10.7191/jeslib.2013.1041 [Accessed on: 30 January 2015].
33. Guy, M. 2013. RDM Training for Librarians'. *DCC RDM Services case studies.* [Online] Available at: http://www.dcc.ac.uk/resources/developing-rdm-services [Accessed on: 21 November 2015].
34. Heidorn, P.B. 2011. The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration,* 51: 662-672. [Online] Available at: http://0-citeseerx.ist.psu.edu.innopac.up.ac.za/viewdoc/download?doi=10.1.1.463.1416&rep=rep1&type=pdf [Accessed on: 01 November 2015].
35. Higgins, S. 2012. The Lifecycle of Data Management. In: Pryor, G. (ed) *Managing Research Data*. London, UK: Facet Publishing.
36. Higman, R. & Pinfield, S. 2015, "Research data management and openness: The role of data sharing in developing institutional policies and practices", *Program,* 49(4): 364-381.
37. Hofstee, E. 2006. *Constructing a good dissertation. A practical guide to finishing a masters, MBA or PhD on schedule.* Johannesburg, SA: EPE .
38. Jaguszewski, J. and Williams, K. 2013. New roles for new times: Transforming liaison roles in research libraries. [Online] Available at: http://www.arl.org/component/content/article/6/2893 [Accessed on: 22 July 2015].
39. Jones, S. 2014. Roles of research librarians*. Digital Curation Centre.*
40. Jones, S., Ball, A. and Ekmekcioglu, Ç. 2008. The data audit framework: A first step in the data management challenge. *International Journal of Digital Curation,* 3(2):

112-120. [Online] Available at: http://www.ijdc.net/index.php/jidc/article/view/91 [Accessed on: 12 December 2014]

41. Kafel, D., Creamer, A.T. and Martin, E.R. 2014. Building the New England Collaborative Data Management Curriculum. *Journal of eScience Librarianship*, 3(1): 60-66. [Online] Available at: http://dx.doi.org/10.7191/jeslib.2014.1066 [Accessed on: 22 March 2015].
42. Kahn, M., Higgs, R., Davidson, J. and Jones, S. 2014. Research Data Management in South Africa: How We Shape Up. *Australian Academic & Research Libraries,* 45(4): 296-308. [Online] Available at: http://www.tandfonline.com/doi/abs/10.1080/00048623.2014.951910. [Accessed on: 11 December 2014].
43. Kennan, M.A., Williamson, K. and Johanson, G. 2012. Wild data: Collaborative E-Research and University Libraries. *Australian Academic & Research Libraries,* 43(1): 56-79. [Online] Available at: http://www.tandfonline.com/doi/abs/10.1080/00048623.2012.10722254. [Accessed on: 11 December 2014].
44. Kirk, R.E. 2013. *Experimental design: procedures for the behavioral sciences.* 4th Ed. London, UK: Sage Publishers.
45. Kruse, F. and Thestrup, J.B. 2014. Research libraries' new role in research data management, current trends and visions in Denmark. *Liber Quarterly: The Journal of European Research Libraries,* 23(4): 310-335. [Online] Available at: https://liber.library.uu.nl/index.php/lq/article/view/9173 [Accessed on: 11 December 2014].
46. Kumar, R. 2011. *Research methodology: a step-by-step guide for beginners.* 3rd Ed. London, UK: Sage Publishers.
47. Lamar Soutter Library. 2015. New England Collaborative Data Management Curriculum: Modules for Managing Research Data. *University of Massachusetts Medical School* [Online] Available at: http://library.umassmed.edu/necdmc/modules [Accessed on: 22 November 2015].
48. Leedy, P.D. and Ormrod, J.E. 2010, *Practical research: planning and design.* 9th Ed. Boston, Mass: Pearson Education
49. Leedy, P.D. and Ormrod, J.E. 2013. *Practical research: planning and design.* 10th Ed. Boston, Mass: Pearson Education.
50. Lyon, L. 2007. Lyon, L. (2007) Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report. *UKOLN*. [Online] Available at: http://0-

opus.bath.ac.uk.innopac.up.ac.za/412/1/dealing_with_data_report-final.pdf [Accessed on: 08 July 2015].

51. Lyon, L. 2012. The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation,* 7(1): 126-138. [Online] Available at: http://www.ijdc.net/index.php/jidc/article/view/210 [Accessed on: 12 December 2014].

52. National Research Foundation (NRF). 2015. *Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research.* [Online] Available at: http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded [Accessed on: 11 November 2015].

53. O'Malley, D.L. 2014. Gaining Traction in Research Data Management Support: A Case Study.", *Journal of eScience Librarianship,* 3(1): 74-79. Available at: http://dx.doi.org/10.7191/jeslib.2014.1059. [Accessed on: 15 January 2015].
54. Osswald, A. and Strathmann, S. 2012. The role of libraries in curation and preservation of research data in Germany: Findings of a survey. *78th IFLA General Conference and Assembly.* [Online] Available at*:* http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-osswald-en.pdf. [Accessed on: 06 August 2015].
55. Pickard, A.J. 2013. *Research methods in information.* 2nd Ed. Chicago, Ill: Neal-Schuman.
56. Pinfield, S., Cox, A.M. and Smith, J. 2014. Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE,* 9(12): 1-28. [Online] Available at: http://www.ariadne.ac.uk/issue70/cox-et-al [Accessed on: 11 December 2014].
57. Proctor, R., Halfpenny, P. and Voss, A. 2012. Research Data Management: opportunities and challenges for HEIs. In: Pryor, G. (ed.) *Managing Research Data.* London, UK: Facet Publishing.
58. Pryor, G. 2012. Why manage research data? In: Pryor, G. (ed.) *Managing Research Data.* London, UK: Facet Publishing.
59. Pryor, G. 2014. Options and approaches to RDM service provision. In: Pryor, G., Jones, S. and Whyte, A. (eds.) *Delivering Research Data Management Services: Fundamentals of good practice.* London, UK: Facet Publishing.
60. Pryor, G. 2014. Who's doing data? A spectrum of roles, responsibilities and competencies. In: Pryor, G., Jones, S. and Whyte, A. (eds.) *Delivering Research Data*

Management Services: Fundamentals of good practice. London, UK: Facet Publishing.

61. Pryor, G. and Donnelly, M. 2009. Skilling up to do data: whose role, whose responsibility, whose career? *International Journal of Digital Curation,* 4(2): 158-170. [Online] Available at: http://www.ijdc.net/index.php/ijdc/article/view/126 [Accessed on: 12 December 2014].
62. Ray, J.M. 2014. Introduction to Research Data Management. In: Ray, J.M. (ed.) *Research Data Management: Practical Strategies for Information Professionals*. Purdue, Ind.: Purdue University Press.
63. Research Data Alliance (RDA). 2015. *23 Things: Libraries for Research Data.* [Online] Available at: https://rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_RDA.pdf [Accessed on: 09 September 2015].
64. Rice, R. 2014. Research Data MANTRA: A Labour of Love. *Journal of eScience Librarianship,* 3(1): 44-51. [Online] Available at: http://dx.doi.org/10.7191/jeslib.2014.1056 [Accessed on: 22 March 2015].
65. Richardson, J., Nolan-Brown, T., Loria, P. and Bradbury, S. 2012. Library Research Support in Queensland: A Survey. *Australian Academic & Research Libraries,* 43(4): 258-277. [Online] Available at: http://www.tandfonline.com/doi/abs/10.1080/00048623.2012.10722287. [Accessed on: 11 December 2014].
66. Staller, K.M. 2010. Qualitative Research. In Salkind, N.J. (ed.) *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage Publications.
67. Swan, A. and Brown, S. 2008. The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs. *Report to the JISC*. [Online] Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.8960&rep=rep1&type=pdf [Accessed on: 06 August 2015].
68. Tammaro, A.M. and Casarosa, V. 2014. Research Data Management in the Curriculum: An Interdisciplinary Approach. *Procedia Computer Science,* 38: 138-142. [Online] Available at: http://dx.doi.org/10.1016/j.procs.2014.10.023 [Accessed on: 01 September 2015].
69. Tenopir, C., Sandusky, R.J., Allard, S. and Birch, B. 2014. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research,* 36(2): 84-90. [Online] Available at: http://dx.doi.org/10.1016/j.lisr.2013.11.003 [12 December 2014].

70. Thompson, C.A., Robertson, W.D. and Greenberg, J. 2014. Where Have All the Scientific Data Gone? LIS Perspective on the Data-At-Risk Predicament. *College & Research Libraries,* 75(6): 842-861. [Online] Available at: http://crl.acrl.org/content/75/6/842.abstract. [Accessed on: 11 February 2015].
71. Tiwari, R. and Chand, B.B. 2014. Open Access to Research Data: Sharing Data for Efficient Research. *Indian Journal of Library & Information Science,* 8(3): 239-246. [Online] Available at: http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=48753192-ae2d-42cd-9336-628070e81633%40sessionmgr198&vid=5&hid=110. [Accessed on: 11 December 2014].
72. Treloar, A. 2014. The Research Data Alliance: globally co-ordinated action against barriers to data publishing and sharing. *Learned Publishing,* 27: 9-13. [Online] Available at: http://www.ingentaconnect.com/content/alpsp/lp/2014/00000027/00000005/art00003. [Accessed on: 11 February 2015].

73. Universities South Africa. 2015. *Welcome to Universities South Africa.* [Online] Available at: http://www.universitiessa.ac.za/ [Accessed on: 11 November 2015].

74. University Libraries. 2015. Research Data Services. *University of Maryland.* [Online] Available at: http://www.lib.umd.edu/data [Accessed on: 19 October 2015].
75. University of Cape Town. 2015a. *UCT LISC offers short course in Research Data Management.* [Online] Available at: http://www.lisc.uct.ac.za/news/uct-lisc-offers-short-course-research-data-management [Accessed on: 22 November 2015].
76. University of Cape Town. 2015b. *Senior Data Scientist.* [Online] Available at: http://www.icts.uct.ac.za/ [Accessed on: 09 June 2015].
77. University of Cape Town. 2015c. *Digital Curation Officer.* [Online] Available at: http://www.lib.uct.ac.za [Accessed on: 09 June 2015].
78. University of East London. *TraD - Training for Data Management at UEL* 2015. [Online] Available at: http://www.uel.ac.uk/trad/ [Accessed on: 06 August 2015].
79. University of Edinburgh. 2014. *Do-It-Yourself Research Data Management Training Kit for Librarians*. [Online] Available at: http://datalib.edina.ac.uk/mantra/libtraining.html [Accessed on: 06 August 2015].

80. University of Oxford. 2015. *About RDM*. [Online] Available at: http://researchdata.ox.ac.uk/home/introduction-to-rdm/ [Accessed on: 11 November 2015].

81. University of Sheffield. *About RDMRose*. 2015. [Online] Available at: http://www.sheffield.ac.uk/is/research/projects/rdmrose [Accessed on: 06 August 2015].
82. University of South Africa (UNISA). 2015. *Research Data Curator.* [Online] Available at: http://www.unisa.ac.za/vacancies_view/html/7324.pdf [Accessed on: 20 October 2015].
83. Van Deventer, M. and Pienaar, H. 2015. Research Data Management in a Developing Country: A Personal Journey. *International Journal of Digital Curation,* 10(2): 33-47. [Online] Available at http://www.ijdc.net/index.php/ijdc/article/view/10.2.33 [Accessed on: 11 October 2015].
84. Van Wyk, J. and Van Der Walt, I. 2014. Going Full Circle: Research Data Management @ University of Pretoria. *eResearch Africa 2014*. [Online] Available at: http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/Going_Full-Circle_RDM_UP_VanWyk_VanderWalt_26Nov_2014.pdf [Accessed on: 21 November 2015].
85. Vlaeminck, S. and Wagner, G.G. 2014. On the Role of Research Data Centres in the Management of Publication-related Research Data. *Liber Quarterly: The Journal of European Research Libraries,* 23(4): 336-357. [Online] Available at: http://liber.library.uu.nl/index.php/lq/article/download/9356/9923. [Accessed on: 11 December 2014].
86. Walters, T. 2009. Data curation program development in US universities: The Georgia Institute of Technology example. *International Journal of Digital Curation,* 4(3): 83-92. [Online] Available at: http://www.ijdc.net/index.php/ijdc/article/view/136 [Accessed on: 11 December 2014].
87. Walters, T. and Skinner, K. 2011. *New roles for times: Digital curation for preservation*. Washington, DC.: Association of Research Libraries.
88. Yin, R.K. 2014. *Case study research: design and methods.* 5th Ed. Los Angeles, CA: Sage Publishers.

APPENDICES

Appendix A: First (before) questionnaire

Skills Required for Research Data Management (RDM) - Before training intervention !

Dear Sir/Madam,

I am Refiloe Matlatse, an M. IT (Masters in Information Technology) student at the University of Pretoria. One of the requirements for the completion of the course is to complete a mini-dissertation. The focus of my mini-dissertation is to look at the knowledge and skills associated with Research Data Management (RDM). The purpose of this exercise is to ascertain whether a specific training intervention has a positive effect on your knowledge of RDM.

Participants would be required to complete a questionnaire to share their understanding of RDM before a training event. They would then be required to undergo training and then to complete a similar questionnaire after the training session. The data collected will be anonymised so that identification will be impossible. Your actual identity will not be recorded.

Completing the questionnaire should not take longer than 10-15 minutes.

* Required

Section 1: Consent

1. Do you consent to participating in this study? *

- ○ Yes
- ○ No

Section 2: Demographic Information

2.1. Please select your age range *

- ○ Under 30
- ○ 31 - 50
- ○ 51 and above

2.2. Please select your institution *

- ○ Academy of Science of South Africa (ASSAf)
- ○ Cape Peninsula University of Technology (CPUT)
- ○ Council for Scientific and Industrial Research (CSIR)
- ○ Human Sciences Research Council (HSRC)
- ○ National Research Foundation (NRF)

- North-West University (NWU)
- University of South Africa (UNISA)
- University of Johannesburg (UJ)
- University of Pretoria (UP)
- University of Venda (UV)
- Other:

2.3. Please indicate your position/ job title at your institution *

2.4. Please indicate your years of experience *

- Under 10 years
- 11 - 20 years
- Over 21 years

Section 3: Knowledge regarding Research Data Management

3.1. Does your institution have a policy and mandate to guide RDM? *

- Yes
- No
- I don't know

3.2. In your opinion what aspects should be addressed by the RDM policy and mandate? *

3.3. Does your institution offer any RDM services? *

- Yes
- No
- I don't know

3.4. Please explain what services should be offered.

3.5. In your opinion - Whose responsibility should it be to drive/ lead the initiative to manage research data? *

- [ ] Researcher or research team
- [ ] Subject Librarians
- [ ] Library Technical Services
- [ ] Research funding bodies e.g. NRF
- [ ] Publishers
- [ ] Research office
- [ ] IT Department
- [ ] Other:

3.6. From your selection above - What type of RDM related services should the leading department offer? *

3.7. On a scale of 1-4 please rate your understanding of RDM. *

	1	2	3	4	
Poor understanding	○	○	○	○	Excellent understanding

3.8. Do you think you have the necessary disciplinary background to offer RDM services/ advice? *

- ○ Yes
- ○ No

3.9. Please explain your answer

3.10. Where do you consider your own knowledge gap to be in RDM? *

Please rate your knowledge of the following RDM related areas.

	1 - Poor	2	3	4 - Excellent
Awareness of RDM	○	○	○	○
Long-term preservation	○	○	○	○
Policy and governance of data	○	○	○	○
Copyright and other IP rights	○	○	○	○
Metadata standards for data	○	○	○	○
Trusted repositories	○	○	○	○
Funders mandates	○	○	○	○
Research process / lifecycle	○	○	○	○

3.11. Where do you consider your own skills gap to be in RDM? *

Please rate your skills in the following RDM related areas.

	1 - Poor	2	3	4 - Excellent
Indexing (adding metadata) skills	○	○	○	○
Online citation/referencing for data	○	○	○	○
Data management planning	○	○	○	○
Preparing data for deposit to repository	○	○	○	○
Data licensing	○	○	○	○
Data appraisal and selection	○	○	○	○
Organising and documenting data	○	○	○	○
Data storage and security	○	○	○	○
Create guides or training materials for researchers	○	○	○	○
Working with data				

repositories	○	○	○	○
Managing data repositories	○	○	○	○

Submit

100%: You made it.

Never submit passwords through Google Forms.

Appendix B: Second (after) questionnaire

Skills Required for Research Data Management (RDM) - After training intervention!

Dear Sir/Madam,

Thank you for completing the pre-event questionnaire to share your understanding of RDM as well as for participating in an RDM training event. You are now required to complete a similar questionnaire to establish whether the training had an impact on your understanding of RDM. The data collected will again be anonymised so that identification will be impossible. Your actual identity should not be recorded.

* Required

Section 2: Demographic Information

2.1. Please select your age range *

- ○ Under 30
- ○ 31-50
- ○ 51 and above
- ○ Other:

2.2. Please select your institution *

- ○ Academy of Science of South Africa (ASSAf)
- ○ Cape Peninsula University of Technology (CPUT)
- ○ Council for Scientific and Industrial Research (CSIR)
- ○ Human Sciences Research Council (HSRC)
- ○ National Research Foundation (NRF)
- ○ North-West University (NWU)
- ○ University of South Africa (UNISA)
- ○ University of Johannesburg (UJ)
- ○ University of Pretoria (UP)
- ○ University of Venda (UV)
- ○ Other:

2.3. Please indicate your position at your institution *

2.4. Please indicate your years of experience *

- ○ Under 10 years

- ○ 11 - 20 years
- ○ Over 21 years

Section 3: Knowledge regarding Research Data Management

3.5. In your opinion - whose responsibility should it be to manage research data? *

- ☐ Researcher or research team
- ☐ Information Specialists
- ☐ Library Technical Services
- ☐ Research funding bodies e.g. NRF
- ☐ Publishers
- ☐ Research office
- ☐ IT Department
- ☐ Other:

3.6. From your selection above - What type of RDM related services should the leading Department offer? *

3.7. On a scale of 1-4 please rate your understanding of RDM. *

	1	2	3	4	
Poor understanding	○	○	○	○	Excellent understanding

3.8. Do you think you have the necessary disciplinary background to offer RDM services/ advice? *

- ○ Yes
- ○ No

3.9. Please explain your answer *

3.10. Where do you consider your own knowledge gap to be in RDM? *

Please rate your knowledge of the following RDM related areas.

	1	2	3	4
Awareness of RDM	○	○	○	○
Long-term preservation	○	○	○	○
Policy and governance of data	○	○	○	○
Copyright and other IP rights	○	○	○	○
Metadata standards for data	○	○	○	○
Trusted repositories	○	○	○	○
Funders mandates	○	○	○	○
Research process / lifecycle	○	○	○	○

3.11. Where do you consider your own skills gap to be in RDM? *

Please rate your skills in the following RDM related areas.

	1	2	3	4
Indexing (adding metadata) skills	○	○	○	○
Online citation/referencing for data	○	○	○	○
Data management planning	○	○	○	○
Preparing data for deposit to repository	○	○	○	○
Data licensing	○	○	○	○
Data appraisal and selection	○	○	○	○
Organising and documenting data	○	○	○	○
Data storage and security	○	○	○	○
Create guides or training materials for researchers	○	○	○	○
Working with data repositories	○	○	○	○
Managing data repositories	○	○	○	○

3.12. Please rate the usefulness of the training *

	1	2	3	4	
Not at all useful	○	○	○	○	Very useful

3.13. What would you have liked to be included in the training but it was not? *

3.14. What did you learn that was really useful? *

3.15. Do you feel confident that you would be able to guide research staff when they have RDM questions? *

3.16. Do you have any other comments or suggestions about any future RDM related workshops? *

Submit

100%: You made it.

Never submit passwords through Google Forms.

Appendix C: Training intervention (RDM workshop)

NeDICC Network of Data & Information Curation Communities

Research Data Management Workshop

11 August 2015

Ulwazi, Knowledge Commons, CSIR, Pretoria

08:00-08:30	**Registration**
08:30-08:45	Welcome and Introduction *NeDICC*
08:45-09:30	RDM: what it's all about, the benefits and drivers *Joy Davidson*
09:30-11:00	The impact that RDM has had on the profile and standing of the library at Purdue Univ (via Skype) *Jim Mullins*
11:00-11:15	**Refreshments**
11:15-11:30	CARDIO quick quiz & discussion (Exercise to assess the RDM readiness at SA organizations present at the workshop.)
11:30-12:00	The state of RDM in South Africa *Elias Makonko*
12:00-12:30	DIRISA and its implications for RDM in South Africa *Anwar Vahed*
12:30-13:30	**Lunch**
13:30-14:00	Useful RDM tools and services – resources you can reuse *Joy*
14:00-15:00	DMPonline – brief talk and demo *Sarah Jones*
15:00-15:15	**Comfort Break**
15:15-16:00	Developing a data roadmap for your institution
16:00-16:15	Closing discussion – *Joy & Sarah*