

**IMPLEMENTING A TRUSTED REPOSITORY TO SUPPORT THE
STORAGE AND PRESERVATION OF RESEARCH DATA AT THE
NATIONAL RESEARCH FOUNDATION (NRF)**

BY

ROSE MANALA

(STUDENT NO. 58540008)

SUBMITTED IN ACCORDANCE WITH THE REQUIREMENTS

FOR THE DEGREE OF

MASTERS IN INFORMATION SCIENCE

in the

DEPARTMENT OF INFORMATION SCIENCE

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: Dr ISABEL SCHELLNACK-KELLY

Declaration

Name: Mpho Rose Manala

Student no: 58540008

Degree: Master of Information Science (98410)

**IMPLEMENTING A TRUSTED REPOSITORY TO SUPPORT THE STORAGE AND
PRESERVATION OF RESEARCH DATA AT THE NATIONAL RESEARCH
FOUNDATION (NRF)**

I declare that the above thesis is my own work and that all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

Signature



A handwritten signature in black ink, appearing to read 'Mpho Rose Manala', is written inside a rectangular box. The signature is cursive and somewhat stylized.

Date

Acknowledgements

I would like to extend my heartfelt appreciation to my mentor and supervisor Dr Isabel Schellnack- Kelly, whose knowledge and dedication are the reason I have made it this far.

I am thankful to my family for the support I received from them. My children Ditheto and Oatile are my motivation for succeeding at anything in my life.

I also want to thank my colleagues at the National Research Foundation who encouraged me not to give up but keep working harder.

Abstract

The purpose of this research was to identify critical factors necessary for the implementation of a trusted repository for managing research data. A research data survey was conducted with 36 Information professionals at the NRF to investigate the state of readiness at the organisation for implementing the repository. According to the survey results a need was identified for storage and preservation of student data, databases and reports managed by the Information professionals. Other research data types that are mostly generated were Websites, Databases and manuals that require secure storage. The study further revealed the need for metadata training in the organisation as an important component for research data preservation. More than half of the Information professionals revealed that their metadata is system generated and there is minimum involvement required from them. Through the review of the existing literature and the NRF data survey, the researcher was able to bring forward recommendations that can be applied for improvement of preservation services. Recommendations made include that professionals should be trained, there should be policy development and review of existing research data within the organisation. Secure storage mediums should be an investment for the organisation based on the research data generated through research activities and managed by the NRF professionals.

Sekhetho

Morero oa lipatlisiso tsena e ne e le ho supa lintlha tsa bohlokoa tse hloka halang bakeng sa ho kenya tšebetsong polokelo e tšepahalang ea taolo ea data ea lipatlisiso. Patlisiso ea data ea lipatlisiso e entsoe le litsebi tse 36 tsa Tlhahisoleseling ho NRF ho etsa lipatlisiso tsa boemo ba ho itokisa mokhatlong. Ho latela liphetho tsa phuputso tlhoko e fumanoe ea ho boloka le ho boloka tlhaiso-leseling ea baithuti, marang-rang le litlaleho tse laoloang ke litsebi tsa Tlhahisoleseling. Mefuta e meng ea data ea lipatlisiso e hlahisoang haholo ke Liwebosaete, Libaka tsa boitsebiso le libuka tse hloka halang polokelo e bolokehileng. Boithuto bo tsoetse pele ho senola tlhoko ea koetliso ea metadata mokhatlong e le karolo ea bohlokoa bakeng sa ho boloka lintlha tsa lipatlisiso. Ba fetang halofo ea litsebi tsa Tlhahisoleseling ba senotse hore metadata ea bona e entsoe ka sistimi mme ho na le karolo e nyane e hloka halang ho bona. Ka

ho hlahloba lingoliloeng tse seng li ntse li le teng le phuputso ea litaba tsa NRF mofuputsi o ile a khona ho tlisa likhothaletso tse ka sebelisoang ho ntlafatsa litšebeliso tsa polokeho. Litlhaliso tse entsoeng e ne e le koetliso bakeng sa litsebi, nts'etsopele ea leano le tlhahlobo ea lintlha tse teng tsa lipatlisiso ka har'a mokhatlo. Lits'ireletso tse bolokehileng li lokela ho tsetela mokhatlong o ipapisitseng le tlhaiso-leseling e entsoeng ka mesebetsi ea lipatlisiso le ho tsamaisoa ke litsebi tsa NRF.

Okufingqiwe

Inhloso yalolu cwaningo kwakuwukuthola izinto ezibucayi ezidingekayo ekufezeni indawo yokugcina ethembekile yokuphatha idatha yocwaningo. Ucwaningo lwedatha yocwaningo lwenziwa nochwepheshe bezokwaziswa abangama-36 kwa-NRF ukuphenya isimo sokulungela kule nhlango. Ngokwemiphumela yocwaningo kwahlonzwa isidingo sokugcinwa kanye nokulondolozwa kwemininingwane yabafundi, imininingwane yolwazi kanye nemibiko ephethwe ngochwepheshe boLwazi. Ezinye izinhlobo zedatha yocwaningo ezikhiqizwa kakhulu ngamaWebhusayithi, amaDathabhe namabhukwana adinga ukugcinwa okuphephile. Ucwaningo luqhubeke lwembula isidingo sokuqeqeshwa kwemethadatha kule nhlango njengengxeny ebalulekile yokulondolozwa kwedatha yocwaningo. Ngaphezu kwengxeny yochwepheshe boLwazi baveze ukuthi imethadatha yabo yenziwa ngohlelo futhi kukhona ukubandakanyeka okuncane okudingekayo kubo. Ngokubuyekizwa kwemibhalo ekhona kanye nenhlolovo yedatha ye-NRF umcwaningi ukwazile ukuletha izincomo ezingasetshenziswa ukwenza ngcono izinsizakalo zokulondolozwa. Izincomo ezenziwe kwaba ukuqeqeshwa kochwepheshe, ukuthuthukiswa kwenqubomgomo nokubuyekizwa kwedatha yocwaningo ekhona ngaphakathi kwenhlango. Izindlela zokugcina ezivikelekile kufanele kube wutshalo-mali lwenhlango olususelwa kwimininingwane yocwaningo eyenziwe ngemisebenzi yocwaningo futhi ephethwe ngabachwepheshe be-NRF.

Keywords

Trusted repositories, Preservation, Research data, Data Management, Metadata

Table of Contents

Declaration	ii
Acknowledgements	iii
Abstract	iv
List of tables	v
List of figures.....	x
List of appendices	xi
List of Abbreviations and Acronyms	xii
Chapter 1: Introduction	1
1.1. Introduction	1
1.2. Background to the study.....	2
1.3. Problem statement	3
1.4. Framework of the study.....	4
1.5. Purpose of the study	4
1.6. Objectives of the study	5
1.7. Research question/Sub questions	5
1.8. Delimitations of the study.....	6
1.9. Research Methodology	6
1.10. Definition of terms.....	6
1.10.1. Research data.....	6
1.10.2. Data preservation	7
1.10.3. Research Data management.....	7
1.10.4. Trusted repository.....	7
1.10.5. Information professional	7
1.11. Expected contribution of the study	8
1.12. Justification of the study.....	8
1.13. Ethical Considerations.....	9
1.13. Outline of chapters	10
1.14 Summary	11
Chapter 2: Literature and conceptual framework	12
2.1. Introduction	12
2.2. Open Archival Information System.....	12

2.3. Research data life cycle.....	18
2.4. Research data storage and preservation	20
2.5. Trusted repositories in data management	21
2.6. Policies and standards.....	23
2.7. Empirical work relevant to the study.....	27
2.7.1. International studies	27
2.7.2. South African studies	28
2.8. Summary	28
Chapter 3 Research Methodology	29
3.1. Introduction	29
3.2. Research approach.....	29
3.2.1. Qualitative approach.....	29
3.2.2. Quantitative approach	31
3.2.3. Mixed methods research	32
3.2.4. Approach for this study.....	33
3.3. Research design	33
3.4. Population of the study.....	34
3.4.1. Sampling Technique	35
3.5. Data Collection methods.....	35
3.5.1. Types of Questions.....	36
3.5.2. Online questionnaires for the study	36
3.5.3. Structure of questions	37
Theme1. Background information	39
Theme 2: Research Data Management	40
Theme 3: Research data storage	40
Theme 4: Research data sharing and privacy issues.....	40
Theme 5: Research data management training.....	41
3.6. Validity and reliability of the instruments	41
3.7. Data Analysis procedures.....	41
3.8. Summary	42
Chapter 4 Data Analysis, Presentation and interpretation.....	43
4.1. Introduction	43
4.2. Background of data collection	43
4.3. Data analysis and responses.....	44

4.5. Summary	56
Chapter 5: Summary of the study	58
5.1. Introduction	58
5.2. Summary of findings and research questions.....	58
5.2.1. What is a trusted repository for research data storage?.....	58
5.2.2. What is research data management?.....	59
5.2.3. What data should be stored in a trusted repository at the NRF?.....	61
5.2.4. What are the issues concerned with research data storage and preservation at the NRF? 61	
5.2.5. What policies must be implemented to support a trusted repository?	62
5.3. Suggestion for future studies	63
5.4. Conclusion.....	63
5.5. Recommendations	64
References	66
Appendices.....	84

List of tables

Table 2.1: Australian National Data Services Elements	24
Table 3.1: Advantages and disadvantages of qualitative research.....	30
Table.3.2: Research themes and questions	37
Table 4.1: Respondents overview	43
Table 4.2: Distribution of respondents by departmental units	434
Table 4.3: Professional position overview	435
Table 4.4: Years of experience	436
Table 4.5: Data types	47
Table 4.6: Storage space requirement.....	50
Table 4.7: Storage term	51
Table 4.8: Daily data storage services.....	51
Table 4.9: Long-term data storage services.....	52
Table 4.10: Storage and preservation needs	53
Table 4.11: External repositories for data storage	53
Table 4.12 Granting access to data.....	54
Table 4.13: Privacy and confidentiality issues	55
Table 4.14: Data management Training required.....	55
Table 4.15: Areas of data management training	56

List of figures

Figure 2.1: OAIS Functional Model	13
Figure 2.2: DataOne Data life cycle.....	18
Figure 4.1: Individual or Group participation.....	46
Figure 4.2: Storage services	48

List of appendices

Appendix 1: Research Ethical certificate 84
Appendix 2: Data Collection tool: Survey 87
Appendix 3: NRF Granted Permission to conduct a study 91

List of Abbreviations and Acronyms

ANDS- Australian National Data Services

APARSEN - Alliance Permanent Access to the Records of Science in Europe Network

ARC- Australian Research Council

CCSDS - Consultative Committee for Space Data Systems

COAR - Confederation of Open Access Repositories

DAF- Data Asset Framework

DCC – Digital Curation Centre

FAIR- Findable Accessible Interoperable Reusable

HR- Human Resources

ICSU-WDS- International Council of Scientific Unions- World Data Science
ISO- International Organisation for Standardisation

NARA - National Archives and Records Administration

NRF- National Research Foundation

NSF - National Science Foundation

OAIS - Open Archival Information System

OCLC/RLF - Computer Library Centre/Regional Library Facilities

OECD - Economic Co-operation and Development

RISA- Research Infrastructure and System Administration

RLA-OCLC- Research Libraries Group-Online Computer Library Centre

SPP- Strategic Partnership and Platforms

TRAC - Trustworthy Repositories Audit and Certification

UNISA - University of South Africa

WDS- World Data System

Chapter 1: Introduction

1.1. Introduction

A trusted repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future (Research Libraries Group-Online Computer Library Centre (RLG-OCLC) 2007). Research institutions are required to ensure that the digital repositories implemented for research data are compliant with the checklist of trusted repositories (DCC 2011). Trustworthy Repositories Audit and Certification (TRAC) provides research institutions with a checklist on compliance with the Open Archival Information System from Online Computer Library Centre/Regional Library Facilities (OCLC/RLF) and the National Archives and Records Administration (NARA). Research preservation is concerned with applying best practices to best support the archiving of research results to ensure reuse and generation of new knowledge. A good research storage system ensures retention practices and data availability for access, reuse as well as tracking of data through the data life cycle. Electronic data that is archived for preservation should be complete with metadata to describe the record of the data preserved and audit trails linked to the data (The Organisation for Economic Co-operation and Development (OECD) 2015).

It remains a challenge for research organisations to identify a trusted repository that can be implemented to manage the flow of data and also ensure that relevant processes of certification are followed to ensure compliance. The development of trusted repositories in research organisations has been work in progress as there are gaps that still exist in the literature, especially on South African research institutions. In South Africa there is lack of formalised standards and strategies to measure the quality of a trusted repository. The implementation of standards may take several years to be established as it requires organisational skills and certification to ensure compliance with publishing laws that govern research (Giaretta 2009 & Research Library Group-Online Computer Library Centre 2007).

This research identified factors to consider when developing a trusted repository for managing research data. In analysis of research data behaviour, Akers and Doty (2013) point out that storage mediums used by researchers from arts and humanities,

social sciences, medical sciences, and basic sciences disciplines are important factors to consider in research data preservation. The authors mention data banks, data centres and institutional repositories as the databases that are mainly used mediums for data collection, sharing and storage services. This research investigated the types of storage mediums used by the NRF Information professionals and their effectiveness.

1.2. Background to the study

The National Research Foundation promotes and supports research through funding, human resource development and the provision of the necessary research facilities, in order to facilitate the creation of knowledge, innovation and development in all fields of science and technology, including indigenous knowledge. In this way, the NRF contributes to the improvement of the quality of life of all the people of the Republic, as stated in the NRF Act No.23 of 1998 (NRF 2015). “The National Research Foundation Act 23 of 1998 intends: to provide for the promotion of research, both basic and applied, and the extension and transfer of knowledge in the various fields of science and technology and indigenous technology; and for this purpose to provide for the establishment of a National Research Foundation; and to provide for incidental matters.”

In support of open access to research, the Knowledge Resources Unit within the NRF drafted and published an open access statement that promotes the dissemination of NRF funded publications. The NRF statement puts emphasis on research that supports societal needs, achieving overarching policy harmonisation and new innovative publishing models. When engaging in research activities, South African researchers are guided by the Open Access policy to deposit their research in a trusted repository. Following that, all data generated throughout research life cycles should be deposited in an accredited Open Access repository (NRF 2015).

In 2007 the Organisation for Economic Co-operation and Development (OECD) released principles for accessing public funded research data. The principles included openness, flexibility, transparency and legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability and sustainability. The principles were also developed to measure the costs associated with ensuring access against benefits resulting in

managing the data (OECD 2007). The OAIS (Open Archival Information System) Reference Model was developed in 2003 by The Consultative Committee for Space Data Systems (CCSDS) and adopted by the National Science Foundation in 2007 to outline the representation of information and how that links to data objects in research. The model was also used for developing the research data standards such as accreditation of archives which became a necessity in the introduction of repositories for research data (Giaretta 2009).

In 2010 the National Science Foundation (NSF) announced a Data management plan requirement for published data that resulted from grants and developed an open policy to encourage researchers to deposit literature into repositories to promote access. In its policy it stated that the data management plan should be a primary requirement for grant recipients to encourage researchers to deposit data in an open access repository (NSF 2010). The White House Office of Science and Technology Policy requirement for DMPs was announced in March 2013 to accommodate preservation and access to research data resulting from research grants (Faundeen 2017).

The ARC (Australian Research Council) in 2013 adopted the research data management strategy and officially included it in its collection development policy (Donnelly 2014). Donnelly points out research councils in the UK, US, Australia and Europe which were the first to adopt the notion of research data management and included it in their policy development. The World Data System (WDS) was one of the organisations that showed interest in the development of trusted repositories in the year 2011 through the 29th General Assembly of the International Council for Science (WDS Scientific committee 2015). Through that initiative the WDS and the Data Seal of Approval developed the CoreTrustSeal organisation for the certification of trusted repositories. Open Access repositories play a role in research institutions of tracking, managing and providing access to funded research (Confederation of Open Access Repositories (COAR) 2015).

1.3. Problem statement

The absence of a trusted research data repository at the NRF poses challenges as the NRF hosts vast amounts of research data generated yearly from funded students

research projects. Accountability is a requirement on safe data storage and ready access to the results of research projects. The NRF requires a research data repository that will support both storage, access and dissemination of research data. Implementing data repositories and managing large volumes of research data requires standard procedures to avoid risks, provide sustainability and direction for future research (Poole 2016). Furthermore managing digital research outputs can be costly and labour intensive to archive and ensure long-term security and accessibility (Koopman and de Jager 2016). The lack of a trusted research repository at the NRF is an issue as this poses data security threats and data loss due to technology obsolescence. It is the mandate of the NRF to support knowledge generation, knowledge sharing and knowledge creation, therefore it is critical to build a trusted repository for data curation and preservation. This study aimed at investigating the NRF state of readiness in implementing a trusted research data repository. Furthermore, the processes to be followed through the implementation of a trusted repository and identified factors to be considered to facilitate data preservation and access.

1.4. Framework of the study

This study has adopted a conceptual framework through deriving key objectives from various theories in the literature. The researcher derived key objectives necessary for guiding the implementation of a trusted research repository through frameworks such as the Data Asset framework and the Open Archival Information Systems functional model. Other elements used to guide the research study were the Australian National Data Services, Dublin core metadata standards and the Data life cycle to further inform the literature study.

1.5. Purpose of the study

This study investigated principal factors to consider when implementing a trusted repository for the preservation of research data. The researcher identified tools and technologies that the NRF can apply to add value to the field of research data preservation. It is expected that the study conducted in this dissertation makes contribution to the NRF knowledge resources to implement a trusted repository to be

used for research data storage. Research findings of this report have informed the recommendations made by the researcher that can be used at the NRF to implement a trusted research data repository.

1.6. Objectives of the study

The objectives of this study were to;

- I. Identify research data generated within the NRF that needs to be stored in a trusted repository and processes to follow for the storage
- II. Investigate and inform the NRF on factors, issues and models available in the process of implementing a trusted repository for research data
- III. Give direction to the NRF on relevant standards and policies to apply when implementing a trusted research data repository.
- IV. Identify challenges within the NRF and externally that affects the process of implementing a trusted research data repository.

1.7. Research question/Sub questions

The study is aimed at answering this question:

How can the NRF build a trusted repository for storing research data through applying best practices and ensuring compliance?

A critical approach was used to obtain a list of criteria that research organisations can use for building a trusted repository for storing research data. The sub questions to be answered in this study are:

- i) What is research data and research data management?
- ii) What types of research data is generated within the NRF?
- iii) What is a trusted repository for research data storage?

- iv) What are the issues concerned with research data storage and preservation at the NRF?
- v) What policies must be implemented to support a trusted repository?

1.8. Delimitations of the study

The research data life cycle consists of six phases which are creating data, processing, analysing, preserving, giving access and reusing data (Boston University Libraries 2016). This research was limited to the implementation of a repository which forms part of the last three phases of the research data life cycle, preserving, giving access and reusing data. The study of research data is broad as it explores all phases related to the data life cycle. Trusted repositories focus mainly on how data is preserved for later access and usage, therefore for the purpose of focussing the study on the relevant theory the researcher limited the study to the stages of preserving and giving access to data phases.

The NRF has multiple office branches within South Africa. The researcher chose to focus on the Pretoria Corporate office where specialists dealing directly with research data are located.

1.9. Research Methodology

Chapter 3 covers the methodology and research design applied in the study. The chapter consists of the data collection method that was used in the study and provides an outline on the sampling used for the research study. Furthermore, the themes used to design the method were presented in the chapter. The researcher applied a quantitative approach to explore the state of research data management at the NRF. A group of NRF Information professionals in the NRF Pretoria office were approached to participate in a study survey.

1.10. Definition of terms

Defining research terms is used as a basis to add clarity on how the terms fit into the study. The researcher used the following terms to guide the research study:

1.10.1. Research data

Research data can be defined as any factual information generated and collected throughout a research process. This can be from ideas generated, drafts, graphs and

suggested methods, plans and results throughout the research process. Data is one of the main components of research and development (Thomee, Shamma, Friedland, Elizalde, Ni, Poland, Borth and Li 2015).

1.10.2. Data preservation

Data preservation refers to the activities followed to ensure continuous access to digital material (International Federation of Data Organizations 2018). Moore (2008:64) describes preservation as incorporating new technologies while conserving preservation properties such as authenticity, integrity, and chain of custody.

1.10.3. Research Data management

With the volume of data generated and produced throughout research the importance of data has led to the process of research data management. Research data management is defined as the process of overseeing data that is being generated during a research project (Thoegersen 2014). This is the application of methods such as planning, creating, storing, curating and making data accessible. For data to be effectively managed it has to be stored in trusted repositories to ensure usability over time.

1.10.4. Trusted repository

“A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.” (Research Libraries Group - Online Computer Library Centre (RLG-OCLC) 2007). Furthermore, a trusted repository can be defined as a digital repository that is compliant with the Research Libraries Group/ National Archives and Records Administration (RLG/NARA) certification checklist. The certification checklist outlines a set of policies that establish the attributes of a trusted digital repository based on the Open Archival Information System (OAIS) reference model (RLG-OLC 2007).

1.10.5. Information professional

An Information professional as used in this research study is defined as someone who collects, processes, organises, stores and makes information accessible to the users in printed or digital format. Marchionini and Moran (2012) define Information

professionals as specialists who generate and collect information and form part of research, data and design tea. Marchionini and Moran (2012) associate the term Information Professional with professions such as “information architect, data analyst, database administrator, web developer, ontologist, usability engineer, social media strategist, data curator, chief information officer, v and of course librarian, archivist, and museum curator”.

1.11. Expected contribution of the study

This study aimed at contributing to the NRF knowledge resources department in the implementation of a trusted repository for research data management. Current trends and issues were identified in the literature to effectively guide the NRF in ensuring that there is adherence to policies and standards for data management. With the growing rate of research data generated, research organisations have a challenge of effectively managing data throughout its life cycle. The study outlined the important factors to be considered in the process of research data management where one of the main components is establishing a trusted repository for storage and preservation. Furthermore, this study aimed to provide guiding tools to the NRF in the starting phase of determining the storage of research publications and data. The NRF being one of the leading research funders in Africa is in the investigation phase of implementing research data management and being compliant with national policies. Furthermore the researcher looked at what other research institutions are doing in terms of establishing a trusted repository for managing research by pointing out policies and strategies in place.

1.12. Justification of the study

This study was deemed necessary is there is a gap in the existing literature relating to trusted repositories. Furthermore it is within need for the NRF to adhere to rules and regulations that govern the use of the state public funds by reporting back data that is generated through funded research projects. This is in return proof to the state that there is value for money invested in research. By preserving and conserving research data through a trusted repository, this in return means the South African state can benefit over time by generating new theories from produced research.

1.13. Ethical Considerations

In research projects, ethical considerations are defined as aspects to follow and consider when the research study consists of certain participants. The participants should be protected in terms of confidentiality, privacy and any personal data that can affect their lives (Fouka and Mantzourou 2011). Resnik (2015) clarifies the process of ethical considerations as distinguishing between right and wrong when conducting research. Ethics are guided by the type of study conducted such as theology, psychology or sociology and the researcher should follow ethics at every stage of the research project as it is an integral part of investigations (Resnik 2015 & Parveen and Showkat 2017). The research committee at the University of Gloucestershire (2018) pointed out a list of ethical aspects that should be considered through dealing with research participants. These are:

- The collection, use, and interpretation of research data
- Methods for reporting and reviewing research plans or findings
- Relationships among researchers
- Relationships between researchers and those that will be affected by their research
- Means for responding to misunderstandings, disputes, or misconduct
- Options for promoting ethical conduct in research (University of Gloucestershire 2018)

Before conducting this study, approval was obtained from the NRF Knowledge Resources Management and the NRF Human Resources Management since these are requirements of the organisation when conducting a study. The researcher also took into consideration the UNISA ethics policy, research policy and the IP policy and as such an application form was submitted to the University of South Africa (UNISA) Ethics Review Committee in order for this research to be compliant. The researcher complied with the NRF ethics policy by applying for consent with NRF management to conduct a survey with the NRF staff members as required. The questions formulated

in the survey were only limited to the data that can be collected from the NRF Information professionals. The researcher was limited to focus the questionnaire on relevant data and to the extent where NRF Information Professionals are involved.

1.13. Outline of chapters

Chapter 1: Introduction and background

This chapter presented the background of the study, the objectives, as well as how the study will contribute to the NRF organisation. The researcher obtained and received approval from the Director IT& Knowledge Resources and NRF HR office to peruse the study at the NRF. Furthermore in this chapter the researcher adhered to ethical concerns which include confidentiality, informed consent for participants of the study and plagiarism rules which ensure that copyright is taken cognisance of in the research.

Chapter 2: Literature review

This chapter reviews relevant literature. The data management cycle is outlined in this chapter guiding the theory of the study and how a trusted repository fits into the overall process. The conceptual framework is also outlined in this chapter as well as the theory that the researcher adopted for the study.

Chapter 3: Research methodology

This chapter outlines how data was collected to support the study. The research methods are outlined in relation to the selected population for data collection. The researcher presents the selected method of research to be used for the study.

Chapter 4: Data analysis

The data collected in the research study is analysed, interpreted and summarised in this chapter. All results and findings are presented in detail in the chapter.

Chapter 5: Summary, recommendations and conclusion

Recommendations are made from the results of the data analysis. A conclusion of the overall study is also provided.

1.14 Summary

This chapter outlines an introduction to the research study that focussed on the implementation of a trusted research repository at the National Research foundation. The researcher provided an overview of the study by providing research objectives and questions that guided the literature review. A mixed research approach was chosen for this study and the researcher provided a population to be used for the purpose of the research.

Chapter 2: Literature and conceptual framework

2.1. Introduction

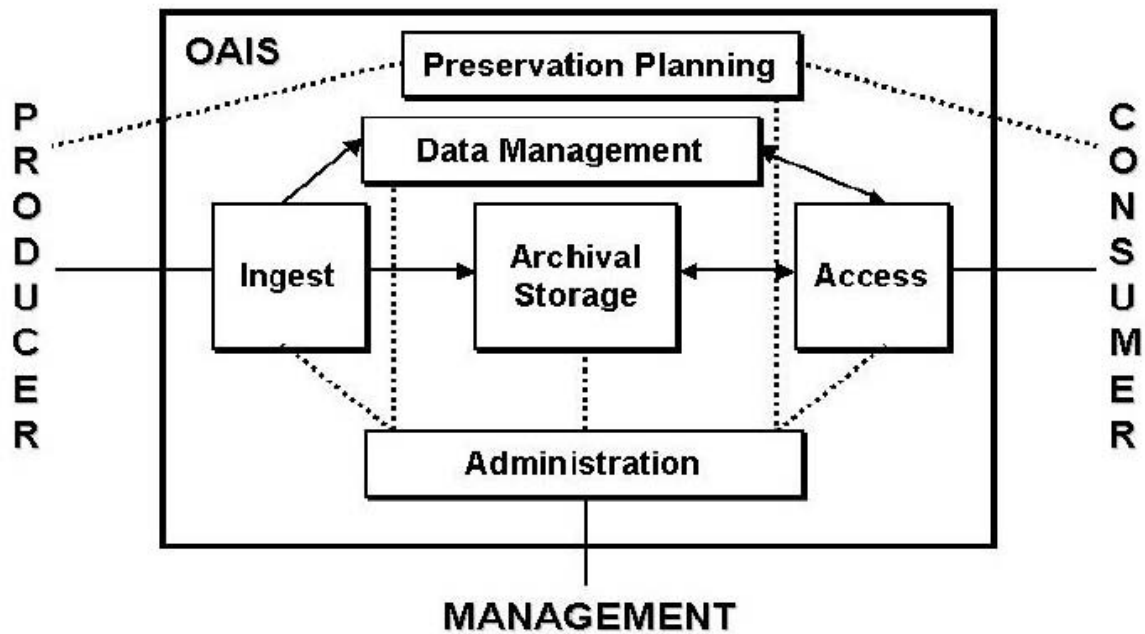
This section focussed on the discussion and analysis of the available literature on trusted repositories. A literature review covers content that is relevant to the study produced by other researchers in that specific field (Howard 2014). The researcher consulted relevant sources and presented an outline of existing arguments from researchers in the field of data storage and trusted repositories. Gaps in the literature were identified and presented in this chapter. The researcher is aware of the broad existing information sources but limited the study to the NRF library catalogue for books, Google scholar for articles and presentations as well as the University of South Africa (UNISA) institutional repository for thesis and dissertation to find relevant literature to the study. The Journal of the American Society for Information Science and The University of Pretoria Institutional repository are other sources that were consulted for relevant literature.

The researcher identified key concepts from various theories that exist in the literature. These concepts are derived from models and systems applied in research organisations to guide the data management processes. The researcher adopted the Open Archival Information System and the Data life cycle as the most commonly used models to guide the implementation of trusted repositories. These models will be discussed in detail in this chapter. Furthermore other important components of trusted repositories will be discussed being technology for storage and preservation, policies and standards to guide the implementation process of a data repositories.

2.2. Open Archival Information System

Downs and Chen (2013) derived metrics that research institutions can improve on to ensure that their repositories comply with the ISO 16363 (International standard for Audit and Certification of Trustworthy Digital Repositories) a standard that sets out rules for archives according to OAIS (Open Archival Information System).

Figure 2.1: OAIS Functional Model



The researcher adopted the OAIS model above to guide and conceptualise the investigation of the NRF state of readiness on implanting a research data repository. The model also serves as a guiding tool to conceptualise objectives and research questions for the study. The OAIS model above defines six core functions involved in the preservation of research data, Lavoie (2014). The functions outlined are:

1. Ingest

In this process research data is ingested by the producer into a data storage system. The producer is responsible for selecting and preparing which data should be ingested in the archival system.

2. Archival Storage

This portion of the system manages long term preservation of data that is ingested in the archival system. The archival storage portion of the system ensures that data is prepared, checked for errors and ready for long term storage.

3. Data management

The data management portion in the archival system ensures that metadata is supported and applied to the information to assist in recovery and in accessing the data. In this component, search queries and retrieval technique functions are supported.

4. Preservation planning

The model preservation and planning is concerned with the strategy put in place for long term preservation of the data ingested in the system. This function ensures viability of the system in an evolving technology environment.

5. Access

This function provides user interface that processes data queries and sends them to the system for processing and returning data requested. This is referred to as the primary function of the OAIS model as it ensures delivery of data to the users.

6. Administration

This involves the day to day operations of the OAIS in terms of managing the interaction with producers and users. This function also supports the management of the OAIS system in relation to policies and standards that support the system.

The OAIS (Open Archival Information System) projects metrics that are applicable for the implementation of research repositories. The metrics are:

1. Governance and Organisational Viability
 - Mission statement and policies - emphasize commitment to continuing stewardship and preservation of scientific data and services.
 - Plans for transferring data, operations, responsibilities, and authority to another entity in case of an unforeseen event
 - Preservation plans are to include details of new procedures as they are adopted.
2. Organisational Structure and Staffing
 - Data stewardship training to be completed by new staff and periodically by experienced staff, which includes Open Archival Information Systems (OAIS) standards and terms.

3. Procedural Accountability and Preservation Policy Framework
 - Processes to define designated community for each Archival Information Package (AIP) during data development and data dissemination planning.
4. Ingest: Acquisition of Content
 - Procedures for recording all inventory, verification, and maintenance activities performed on objects and collections.
5. Ingest: Creation of the Archival Information Package
 - Procedures for testing and improving the understandability of each AIP for the designated community
 - Procedures for recording the provenance of activities completed during data development and dissemination.
6. Preservation Planning
 - Procedures to identify, record, and maintain information on software dependencies for each file received.
7. Archival Information Package Preservation
 - Procedures to verify the integrity of digital objects and files.
8. Technical Infrastructure Risk Management
 - Risk management plans to include an organisational risk register containing tracked risk mitigation schedules
 - Procedures to separate circulation copies of AIPs from archival copies (Downs and Chen 2013).

The researcher identified a framework relevant to the study being the Data Asset Framework. Alexogiannopoulos, McKenney and Pickton (2010:9) define the Data Asset Framework (DAF) as a methodology for assessing data management in the organisation. This framework consists of four steps:

Stage 1: planning, defining the purpose and scope of the survey and conducting preliminary research.

Stage 2: identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis.

Stage 3: Considering information life cycle in order to understand researchers' workflows and identify weaknesses in data creation and curation practices.

Stage 4: pulling together the information collected and provides recommendations for improving data management.

The researcher used this framework to aid in the investigation on research data preservation and trusted repositories. Similar to the stages in the Data asset framework, the researcher adopted the steps and followed them when designing the survey for the investigation at NRF.

In the review of existing theories the following key concepts were identified to guide the research study:

2.2.1. Trusted repositories

Altman (2006:1) points out that repositories are meant to store and preserve information generated within organisations. Furthermore, a digital repository promotes access to information and ensures long term security. Royster (2007) confirms in his study that digital repositories are systems implemented by organisations to store and manage organisational information. Royster asserts that research data repositories facilitate access to the content and ensure preservation and maintenance thereof (Van der Merwe, Van Deventer and Patterton 2015). Thibodeau (2007) states that trusted repositories should be registered with professional bodies that support open access of information and records. Implementing a trusted data repository can eliminate issues such as data loss and data access in long-term preservation. The emergence of institutional repositories (IR) is the demand for researchers to ensure that a trusted system is put in place for data generated throughout the research life cycle. Furthermore, it is important to ensure that in the process of identifying a trusted repository, organisational resources and objectives are reviewed and considered (Giarreta 2009). Dobratz, Schoger and Strathmann (2007) assert that a trusted

repository is characterised by accessibility, long-term preservation, security and reliability. Thus, it is important to ensure that research data is stored in highly trusted repositories for long term preservation. Researchers need to be assured that data generated through their research projects is archived with secure technology to ensure longevity and accessibility.

2.2.2. Repository workflows

Mayer, Proell and Rauber (2012:102) state that workflow processes involved when generating data have to be well documented and stored in trusted repositories. Mayer, Proell and Rauber (2012) emphasise that scientific workflow systems enable researchers to share and exchange data. According to the Alliance Permanent Access to the Records of Science in Europe Network (APARSEN) (2011), data storage plays a critical role in the workflow of research data preservation. Furthermore, data storage is an important strategy, as digital file storage can be unreliable and physical file formats can become obsolete. APARSEN (2011) puts emphasis on back-up, storage, security and encryption of data in a trusted repository.

2.2.3. Data loss

Loss of data due to hardware failure, software faults, viruses and power failures can have a very large effect on organisations. Therefore, organisations need to invest in a back-up strategy for their data. When building a data storage system file sizes, file names and restrictions have to be set as high priorities. Data storage systems must be integrated in a secure environment to avoid data loss when sharing files (Chervenak, Foster, Kesselman, Salisbury and Tuecke 2000:191). Thibodeau (2007) points out that the nature of the information object to be stored, its preservation and dissemination are required for a successful and trusted repository. Thibodeau (2007) and Thurston (2015) imply that the success of a trusted repository will be determined by the accessibility of objects, dissemination and long-term preservation of records.

2.2.4. Research data management roles

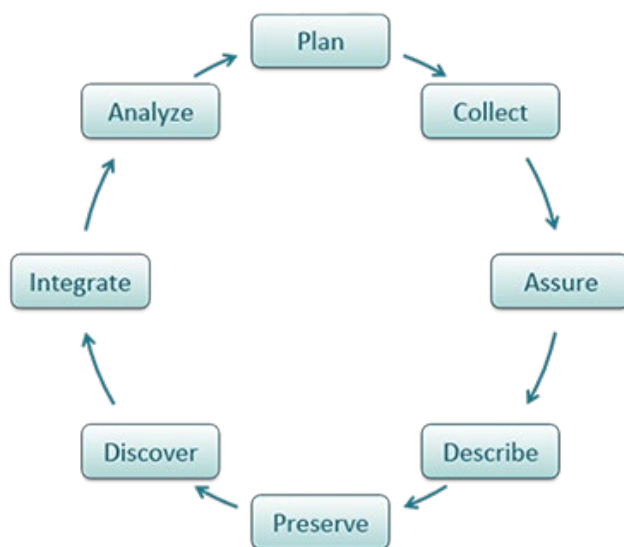
The role of research data management in many organisations has been assigned to librarians as they are known to have expertise in information standards and organisational skills, setting up file structures and their knowledge in collection management (Burnett 2013). Metadata skills are a requirement in storing research

data and have made information professionals vital role players in managing research data preservation. The NRF research facilities are assigned relevant role players who will best implement services on data archiving and preservation (NRF 2015). Molloy and Snow (2012) point out that skills required in research data management are knowledge of the research data, differentiating between research data and published data, organising research data, data sharing, preservation and curation.

2.3. Research data life cycle

The researcher adopted this theme as part of the literature to clarify how data repositories fit in the data life cycle. The main function of a data repository as defined in this research is to facilitate storage, access and preservation of data. Therefore, organisations planning to implement a data repository should evaluate all phases of the data life cycle. Most data generating organisations are currently adopting the research data life cycle to have direction in implementing research data management services. The research data life cycle gives a better overview of the management of data from production to usage (Ball 2014). Alexogiannopoulos, McKenney and Pickton (2010:9) emphasise the importance of an information life cycle in the research data cycle as this defines the behaviour of authors in creation and curation.

Figure 2.2: DataOne Data life cycle



- Plan: description of the data that will be compiled, and how the data will be managed and made accessible throughout its lifetime
- Collect: observations are made either by hand or with sensors or other instruments and the data are placed into digital form
- Assure: the quality of the data is assured through checks and inspections
- Describe: data are accurately and thoroughly described using the appropriate metadata standards
- Preserve: data are submitted to an appropriate long-term archive (i.e. data centre)
- Discover: potentially useful data are located and obtained, along with the relevant information about the data (metadata)
- Integrate: data from disparate sources are combined to form one homogeneous set of data that can be readily analysed
- Analyse: data are analysed (DataOne 2019).

Crane and Chadwick (2017) discuss the stages in the UK Data archive data life cycle and how those can be used as a strategy for implementing research data management. The stages pointed out are creating, processing, analysing, preserving, giving access to data and re-using data. Efficient application of principles and policies throughout the stages of the data lifecycle is required to ensure that data is well managed throughout its life cycle. According to Donnelly (2014), the widely recognised benefits of research data management are that it:

- Leads to faster, more efficient research
- Supports easier exchange between commercial and non-profit research sectors
- Stimulates innovation
- Leads to better use of research infrastructures

Carlson (2014) (as cited in Wissik and Ďurčo (2015:96)) points out three types of data research cycle models: individual-based life cycle models, organisation-based models and community-based models. Furthermore, each model is specific for the intended purpose through drafting data management plans and provides stages that research organisations can use to give clarity to researchers on data management activities (Wissik and Ďurčo 2015: 96-97). The models points out checklists to be used by

research organisations when processing data and providing service for depositing data into a repository.

2.4. Research data storage and preservation

A challenge remains in research organisations to ensure research data storage, backup and security and to minimise data loss (Deards 2013:53 & Freiman, Ward, Jones, Molloy and Snow 2010). Akers and Doty (2013) & Giaretta (2009) assert that finding high quality infrastructure also remains a challenge when identifying technologies for implementing a trusted repository. Developing a trusted repository requires an organisation to identify the relevant technology to ensure quality preservation for future access and security of research data. Digital preservation is defined as the “series of managed activities necessary to ensure continued access to digital materials for as long as necessary” (Digital Preservation Coalition 2019). A digital archive is referred to as a repository as it serves to archive research data. Corti, Van Den Eynden, Bishop and Morgan-Brett (2011) assert that research data storage should guarantee data security and long-term access. Furthermore, the purpose of preserving research data should be to make it sharable to support new scientific research. Preservation ensures that materials are stored in a repository and make data accessible to the designated community over long term periods (RLG-OCLC 2002). Furthermore, there should be regular checks on authenticity and integrity as part of repository administration. Bailey (2008) (as cited in Gladney 2012:210) states that preservation can be defined as the subset of digital curation. Furthermore, preservation should ensure a saved copy of every preserved document, access and use of documents, trustworthiness, no technical complexity and automatic procedures to reduce human effort (Gladney 2002:207).

Thurston (2015:33) asserts that a trusted repository is implemented to ensure that data is archived using standardised digital preservation models. Furthermore, good data is derived from the quality of record management systems in an organisation. Repositories identified have to be interoperable to existing records management systems in the organisations to ensure ease of identification and access. The significance of digital trusted repositories also plays a key role in maintaining the quality of the data and preventing data loss. Thurston (2015) further mentions that the

basics to the development and growth within a nation are the quality of data gathered and the ability to retrieve it from trusted storage repositories. Other issues pointed out are accuracy of the context that determines how the data will be used and by whom it will be accessed. It also includes the challenges in the evolving digital systems as this can lead to data loss if ignored. Identifying a trusted repository is vital as this will ensure the effectiveness of the data life cycle. Researchers need to be assured that their research data is securely stored in high quality repositories.

2.5. Trusted repositories in data management

Rankin (2005), Crow (2002) and Lynch (2003) define an institutional repository as sets of services for storing and making available an institution's intellectual output. Harmsen (2008) asserts that data repositories are responsible for long-term preservation and distribution of research data. The Consultative Committee for Space Data Systems (2011) provides recommendations for space data systems practices that can be used to obtain certification of trustworthy digital repositories. The International Council of Scientific Unions- World Data Science (ICSU-WDS) (2019) developed criteria to be used to assess repositories for trustworthiness. The components focussed on are "authenticity, integrity, confidentiality, and availability of data and services".

Poole (2016) asserts that organisational infrastructure such as archives and institutional repositories are necessary and required for data storage and sharing. Furthermore, repositories remain an important component of digital curation and so data generating agencies should not be neglected in research data management. A trusted repository is evaluated and should meet the guidelines as set out by Data Archiving and Network Services. These guidelines are established to ensure that the data repository complies with integrity, authenticity, legal regulations and quality management. The guidelines further emphasise long-term durable archiving and that the repository agrees with organisational policies and codes of conduct in higher education (RLG-OCLC 2002). The Consultative Committee for Space Data Systems (CCSDS) (2011) further points out that a trusted repository should be able to identify threats, maintenance and monitoring schedules to ensure successful data preservation.

Faundeen (2017:10) categorises five guidelines to be used by research organisations when implementing a data repository namely “storage and geographic location, file fixity and data integrity, information security, metadata, and file formats”. Faundeen outlines the elements that support the above guidelines, which are data access, maintenance, data creating and curation, repository staff, data integrity and authenticity. Furthermore, other elements outlined are long term preservation and the ability to disseminate data to the intended community. Lyle (2019) outlines that the CORETRUSTSEAL trusted repository certification as one of the tools research intuitions can use to refer to guidelines on implementing trusted repositories for research data. Lyle mentions transparency processes and procedures and community standards as important factors to consider when applying for assessing organisations for trusted repositories. Research organisations in the process of implementing a trusted repository are required to register to obtain membership and certification that will assist in determining and adhering to the characteristics of trusted repositories (Dillo and de Leeuw 2018). Dillo and de Leeuw points out that CORETRUSTSEAL offers research organisations a Core Trustworthy Data Repositories Requirements catalogue and procedures for obtaining certification.

A trusted repository should be compliant with the Reference Model for an Open Archival Information System. The Open Archival Information System supports the functional model that provides storage and access as well as the information model that supports long-term maintenance and access for future use of research data (Pellegrino 2014). When implementing a trusted repository, certain guidelines are followed which support the depositors in terms of long term maintenance of digital information and provide a system that supports the long-term viability of digital information. The repository should also have policies, practices and performance that can be audited to ensure the responsibility it holds of trustworthiness and sustainability (RLG-OCLC 2002).

There are three components in research data management: the data producer, the data repository and the data consumer. Research data is deposited in the repository based on the file formats that are recommended by the data repository. When depositing material in the data repository, there is metadata required to ensure the description of research data (Harmsen 2008). Bezuidenhout and Macanda (2014)

describe research data management as the process of digitisation and preservation of raw research irrespective of the format and location. Corti, Van den Eynden, Bishop and Morgan-Brett (2011) define the process of research data management as selecting, ingesting, curating and providing access to data. Corti, Van den Eynden, Bishop and Morgan-Brett (2011) place emphasis on the storage and access to research data. The definition relates to the study as the researcher aims to identify the suitable storage medium for research data.

2.6. Policies and standards

The lack of research data management policies and practices remains a challenge in research institutions as most institutions are reliant on adopting international standards such as the NSF policy on Data sharing policy (Halbert 2013). Policies will assist research organisations to measure compliance of the trusted repository that will house publications that are a result of public funding (Hodson and Molloy 2014). In 2012 the Consultative Committee for Space Data Systems developed ISO 16363 (International standard for Audit and Certification of Trustworthy Digital Repositories) when the need to have a standard for digital repositories was identified. This standard was developed as there were limited guidelines to organisations for selecting trusted digital repositories (Downs 2017).

Organisations use the ISO 16363 standard for assessing the viability of processes when implementing trusted repositories. This standard supports the theoretical nature of OAIS (Open Archival Information System), as it gives direction for compliance when archiving digital objects (Houghton 2015). According to Houghton (2015:3), ISO 16363 covers three main areas and has 105 criteria used as guidelines for assessing repositories. There areas are:

- **Organisational infrastructure**
-including governance, organisational structure and viability, staffing, accountability, policies, financial sustainability and legal issues
- **Digital object management**
-covering acquisition and ingest of content, preservation planning and procedures, information management and access
- **Infrastructure and security risk management**

-covering technical infrastructure and security issues

Ball (2010:40) points out the OpenDOAR Policies as one of the tools adopted by research institutions in the process of developing open access repositories. Ball emphasises that the tool assists in defining metadata, handling data, content, submission and preservation policies in open access repositories. One of the adopted approaches is the process of developing a research data framework that will guide in developing a policy for research data services (Alexogiannopoulos, McKenney and Pickton 2010:5).

The Australian National Data Services (ANDS) provides elements that are considered when developing a research data framework and they are policies and procedures, IT infrastructure; support services, managing metadata, managing research data (ANDS 2018). One of the important elements to consider is metadata. Metadata is defined as data that used to describe data and is applied to ensure that the data curated is well understood by users (Dublin core Metadata Initiative 2014). Furthermore, other benefits of applying metadata are data discovery, avoiding duplicated efforts when curating data, visibility and sharing of data (Jones 2011). Jones also points out citation and improved research as other benefits that researchers who use data can benefit from metadata.

The researcher considered the use of these elements to structure the research study. Some of the research councils that have adopted the ANDS research data guidelines are the Australian Research Council (ARC), and trusted repositories should be implemented in line with publishing standards and open access standards (Hugo 2015). The National Science Foundation implemented a data sharing policy in 2011 to encourage the dissemination and sharing of research results by all researchers who are NSF grant recipients. These include primary data, samples, physical collections and research created or gathered in the course of work under NSF grants (NSF 2010). Furthermore, the NSF data sharing policy emphasises making research results open to encourage research collaborations and knowledge transfer.

The elements of the ANDS are presented as follows (ANDS 2018):

Table 2. 1: Australian National Data Services Elements

Elements	Level 1 Initial	Level 2 Development	Level 3 Defined	Level 4 Managed	Level 5 Optimised
	Process is disorganised & ad hoc	Process is under development	Process is standardised, communicated	Process is managed, measured	Focus is on continuous improvement
Institutional policies & procedures	Policies & procedures may be undeveloped, not up to date, and/or inconsistent.	Policies & procedures are developed & harmonised.	Policies & procedures are promulgated & absorbed into behaviours.	Policies & procedures accepted as part of the culture & subject to audit.	Policies & procedures are subject to review & improvement.
IT infrastructure	IT infrastructure provision is patchy, disorganised & poorly publicised.	Funds are invested in technology & skills. Responsibilities are defined. Processes are established, defined & documented.	Management shows active support. Facilities are well defined & communicated, standardised & integrated.	Funding adapted to need. Management is actively engaged. Documentation kept up to date.	Concerted efforts to maintain, update & publicise infrastructure. Metrics & feedback used to optimise services.
Support services	Training is ad hoc, curation & preservation services are disorganised, data management planning is unsupported & other services inconsistent	Investment in skills. Services identified & staffed. Responsibilities are defined. Documentation & training developed.	Active participation in training with widespread availability of support services.	Widespread take up of services. Curation & preservation acknowledged as critical to the institutional mission.	Customer feedback used extensively to update & improve services.

	& poorly publicised				
Managing metadata	Metadata management is adhoc, chaotic & understood by only a few and without established standards.	Responsibilities are defined & skills developed. Processes are established & documented. Metadata applied to key datasets & shared externally.	Processes are standardised & integrated. Metadata created for new datasets & shared externally, to ensure data is findable and accessible.	Metadata quality metrics are collected. All datasets described in machine readable format & metadata shared. Metadata aligns with the FAIR data principles	Continuous improvement applied to processes & capabilities.
Managing Research Data	Data is stored in ad-hoc facilities. Only data custodians know where the data is stored and accessible only to the researcher or small group of researchers,	Standard formats are not applied and the potential for reusability is limited. Institutional data storage facilities are being developed. Data standards established	Well defined data storage facilities are established. Data managed using standard open formats widely used. Some connectivity between systems permitting limited reuse and sharing	Data routinely stored in established repositories in machine-readable formats using open standards in established facilities according to FAIR principles.	Continual improvements to maintain update & publicise infrastructure. Metrics & feedback used to optimise services.

The NRF Open access statement emphasises that a trusted repository should be developed according to data storage requirements and governance hence research institutions must ensure that there is a data management plan and policy in place (NRF 2015). Furthermore, research institutions are required to have open access deposits for the research data produced during scientific experiments. There is a requirement

for certification of digital repositories and a register of trusted repositories that research organisations must comply to.

In the RLG-OCLC (2002) report, it is pointed out that trusted repositories must accept conventions and standards to ensure that there is ongoing management of research data, access and secure environments for storage. Furthermore, there should be policies and practices that can be audited and measured to ensure compliance with relevant standards. The research aims to investigate relevant standards that the NRF must comply with in the process of implementing a trusted repository for managing research data. Although currently in South Africa there is lack of standards for governing trusted repositories, organisations like the NRF and higher academic institutions are working towards the implementation of South African policies and standards for research data management and trusted repositories (NRF 2015).

2.7. Empirical work relevant to the study

2.7.1. International studies

The researcher reviewed work that is empirical to the study conducted. The researcher is aware of the research data management studies dated from 2001 and 2015, however has chosen to use studies to from 2015 up to 2019 that were conducted globally. A study by Whitmire, Boock and Sutton (2015) focussed on the guidelines by federal agencies to facilitate access to research results from funded projects. The work is more relevant to this research study as the researcher aimed to investigate how the NRF can develop a trusted repository to make funded research results publicly accessible. The study further investigated how research data specialists develop “public access plans” to support research data management within the agency.

A study by Peset et al (2015), reported a study on the curation of research data in Spain through a dataset project survey. The researchers conducted an audit to investigate the status of researchers and their behaviour in managing research data. The study is relevant to this literature study as the researcher focusses on NRF information professionals’ behaviour in research data management practices. A study by Kennan and Markauskaite (2015) presented a study on data management sharing

and practices amongst researchers in publicly funded projects. The study covered issues such as researchers finding safe storage and data sharing. Furthermore the study revealed the importance of data management plans, policies and practices which relates to researchers in publicly funded projects.

2.7.2. South African studies

Coetzer (2018) investigated the process of designing a research data management system for the Hartebeesthoek Radio Astronomy Observatory (HartRAO) in Gauteng. In the study the author focussed on the system design, criteria to be followed, characteristics and components of the research data management system. Further to that, the infrastructure architecture and data organisation. This South African study contributed to the knowledge by highlighting the challenges to expect in the process of designing a research data management system. Patterton, Bothma and Van Deventer (2019) piloted a South African study on the readiness on South Africa research institutions to implement research data management. The study focussed on investigating the needs and habits of researchers in research data management. The 2 above mentioned studies were selected as they were relevant to this literature and assisted the researcher in designing a research methodology for this study.

2.8. Summary

The literature review was conceptualised on the themes that emerged in the implementation of trusted repositories and research objectives that had to be addressed in the study. Themes that were covered in this chapter are the OAIS model, the Data lifecycle. These themes are followed by research organisations to ensure that repositories are compliant with international and national standards for the preservation of research data. The researcher also discussed the challenges that affect implementation of trusted data repositories being lack of technology for storage and preservation of research data. Furthermore, the concept policies and standards was discussed as it forms part of the integral process of implementing research data management. The researcher also looked at an overview of the NRF policies and procedures that will guide the implementation of the NRF trusted repository.

Chapter 3 Research Methodology

3.1. Introduction

A research methodology is concerned with steps followed to collect and analyse data in a research study (Scotland 2012). In this study the researcher applied a quantitative approach to investigate research data preservation at the NRF. The aim of applying the methodology was to be objective and focus the study on relevant skills and experience in working with research data management. A researcher used the questionnaire to determine the level of knowledge in research data management amongst Information professionals. The researcher also investigated the options that researchers at the NRF opt for when managing their research data and the level of application of research data management practices. Cibangu (2012) discusses the role of research methodology as a component of knowledge acquisition through observation and experiment against human experience. The researcher obtained all relevant participants by investigating a list of NRF professionals who deal directly with processing research data within the organisation. 46 participants from 10 NRF departmental units were identified from and selected to be a suitable population for the research study.

3.2. Research approach

A research approach is defined by Creswell (2014) as a plan and a procedure with list of steps on data collection, analysis and interpretation. Creswell (2014:31) emphasizes three approaches to analysing scientific research which are quantitative, qualitative and mixed research methods. When the two methodologies are used together in the same study that is referred to as mixed methods approach.

3.2.1. Qualitative approach

Burns and Grove (2011:19) define qualitative research as “a systematic subjective approach used to describe life experiences and situations to give them meaning.” Furthermore, qualitative research is used to gain insight by investigating detailed facts to the research problem. Borbasi and Jackson (2012) assert that qualitative research is concerned with the application of non-statistical methods to study human experiences. Crotty's (as cited in Creswell 2014:35-37) points out that in qualitative research, human beings share their historical and social view of the world and

researchers use open-ended questions to analyse their views. Johnson and Christensen (2004:645-646) assert that qualitative research methods are often used to investigate the objectives of the study and one good example is the use of telephonic interviews for collecting data.

Wallace and Van Fleet (2012) point out that qualitative research is concerned with naturalistic and particular approaches that emphasise the depth of the study rather than quantity within the study. The qualitative method focusses more on the study of social impact on human and individual understanding. The author further points out that where there is uncertainty on which method to apply, the researcher can use mixed methods as a combination. A case study in Baškarada (2014:1) outlines how the qualitative approach is often misunderstood and the wrong application of the method can impact negatively on the research study. In the nature of research, there are misunderstandings that the researcher can make when conducting a study through qualitative methods. The first misunderstanding is where theoretical knowledge is made more important than practical knowledge. Secondly, researchers tend to make general results from individual cases rather than consulting more case studies. Another problem arises where the case study is only suitable for generating hypothesis than testing hypothesis and building theory (Flyvbjerg 2006:221). Lastly, there can be bias towards verifying a study and difficulty in summarising theories based on specific case studies (Wallace and Van Fleet 2012).

Mohajan (2018:19) discusses the advantages and disadvantages of qualitative research as tabled below:

Table 3.1: Advantages and disadvantages of qualitative research

Advantages	Disadvantages
Open-ended questioning reveals new or unanticipated phenomenon, and raises more issues through broad and open-ended inquiry	Difficult to demonstrate the scientific rigor of the data collection exercise.

Includes a diverse and representative cross-section of affected persons.	It is generally open-ended; the participants have more control over the content of the data collected.
It is in-depth analysis of the impact of an emergency.	It does not have a preconceived, finite set of issues to examine.
It is rich and detailed information about affected populations.	Results in data are not objectively verifiable.
It allows researchers to explore the views of homogenous as well as diverse groups of people, hence helping in unpacking these differing perspectives within a community.	Collection of the data can be time consuming and costly.

3.2.2. Quantitative approach

Quantitative research is concerned with the application of scientific methods to analyse a research problem; often the researcher relies on the validity of the research instruments (Borbasi and Jackson 2012). A quantitative research method is characterized by statistical analysis; it uses variables to test theories. The application of techniques such as questionnaire is used in collecting data for quantitative research (Creswell 2014). Cibangu (2012:96) points out that quantitative research is more concerned with formulating predictions from research studies, and researchers use this method to derive the controllability and manipulation of research. Cibangu (2012) emphasises that researchers use methods like questionnaires and samples to minimize the scope of research for better analysis. A quantitative design involves the application of numbers to draw data and analyse findings from research participants where qualitative methods focus more on the depth of research findings. Johnson and Christensen mention one of the strengths of using quantitative methods as being a quick way to obtain and study a large sample of populations.

Ospina (2004), in Mohajan (2018:21), lists the benefits of using quantitative research:

- *It is more reliable and objective.*

- *It often reduces and restructures a complex problem to a limited number of variables.*
- *It looks at relationships between variables and can establish cause and effect in highly controlled circumstances.*
- *It tests theories or hypotheses.*
- *It assumes a sample which is representative of the population.*
- *Subjectivity of researcher in methodology is recognised less.*
- *It is less detailed than qualitative data and may miss a desired response from the participant.*

3.2.3. Mixed methods research

Johnson and Christensen (2004:648) & Venkatesh, Brown and Sullivan (2016: 437) observe that mixed research applies when the researchers use both qualitative and quantitative research methods in one research study. Multiple research methods are applied by most researchers to obtain better data quality and to answer research questions from two different approaches. Researchers apply qualitative methods in order to ensure a detailed understanding of the research problem and use quantitative methods to obtain a general understanding of the research problem (Ngulube and Ngulube 2015).

Furthermore, mixed methods is recognised to be the third major used research approach in conducting research that aims to include multiple viewpoints by using both quantitative and qualitative research methods (Johnson, Onwuegbuzie and Turner 2007:113). Johnson, Onwuegbuzie and Turner (2007) affirm that mixed method research has been receiving new terminology in current research trends i.e. integrative research, multimethod research, triangulated studies and ethnographic residual analysis. In scientific research, the most commonly used research methodologies are qualitative and quantitative; in addition, applying both methods to one study referred to as mixed method is also common (Lichtmann 2014). Lichtmann (2014) is of the view that both methods can be applied to form a new approach to research rather than

be used in opposition. Wallace and Van Fleet (2012: 26) and Cibangu (2012) assert that the two research methods are often used as opposites and compared when they should both be used for different hypothesis in a study.

3.2.4. Approach for this study

The researcher's intended purpose was to gather information on the state of readiness from NRF on implementing a research data repository. Furthermore to investigate the views and knowledge of information professionals with regards to trusted repositories. Based on the intended purpose, the researcher selected the qualitative approach to gather the relevant evidence to support the study. The researcher applied a quantitative approach using a questionnaire attached on this report as appendix 2 to investigate the need for implementing a trusted repository for managing research data at the NRF. The researcher used existing literature on trusted repositories to formulate a questionnaire for the purpose of this investigation. A qualitative approach was selected as it will offer in-depth evidence that allows the researcher to investigate the gaps and challenges at the NRF for implementing a research data repository.

3.3. Research design

The researcher chose to use a survey for investigating the implementation of a trusted repository and has selected the applied quantitative research approach to investigate the topic. The researcher designed a questionnaire to collect data from the selected group of participants at cross sectional time frames using an online questionnaire. A survey can be used to share knowledge and to test assumptions and develop new theories from research results (Gideon 2012). Baškarada (2014:5) defines research design as what links the research questions to the conclusion. Furthermore, research design can be linked to the steps taken through collecting data and analysing it. Denzin and Lincon (as cited in Creswell 2014:11-14) refer to research designs as "strategies of inquiry" that provide guidance in answering research questions. Ngulube and Ngulube (2015) list six methods that derived from different research designs applied in research. These are case study design, motivation research, survey design, operations research, model building and simulation. Furthermore, the study pointed out survey design as the method mostly applied in research then followed by the case

study design. Creswell (2014:41) points out that research designs are types of inquiry that exist in research approaches and they provide direction for procedures in research.

Thomas (2010) further asserts that research design serves as a “master plan” for the research study and points out the research methods to be followed for investigating a study. Edmonds and Kennedy (2012:7) are of the opinion that research design should focus both on logistical and theoretical considerations. Further, research design provides a framework for the researcher on the questions and hypothesis to apply during a research study.

3.4. Population of the study

Creswell (2014) & Hardon, Hodgkin and Fresle (2004:57) define a population as a group of people selected with the intention to be used in a research study. The population to be used for this study are the NRF information professionals as they are primarily responsible for data management in the organisation. The information professionals selected for the study are IT specialists, Information specialists, Liaison officers, Professionals officers, Librarians, and statisticians. Wallace and Van Fleet (2012: 145) describe census as a descriptive study of an entire population. The Australian Bureau of Statistics (ABS) (2014) states that there are two ways to study a population, through a census and taking sample. For the purpose of this study a census was used to identify and contact the NRF information professionals for this research study

As mentioned previously, the researcher selected the NRF Information professionals that are responsible for the day to day management of research data in the organisation as a selected population for the research study. Information professionals were selected with the assistance of the Information Manager at the NRF, by the use of the staff database. This selection was also motivated by the area of work of the professionals as this involves day to day processing of research data. The aim is to collect data from information professionals on their understanding and usage of trusted repositories in data management. The data will also be used to determine underlying issues with regards to identifying a suitable trusted repository.

3.4.1. Sampling Technique

O'Dwyer and Bernauer (2014) point out two methods of sampling which are probability and non-probability sampling. In probability sampling all the elements have a known probability whereas in non-probability sampling elements from participants cannot be established. Creswell (2014) asserts that sampling in research can be random and non-random.

The researcher opted to use a census for the research study for the purpose of collecting data from experienced and skilled professionals in research data management. The selection of this technique was also informed by the nature of the research study which seeks to investigate and gather evidence from experienced professionals. The researcher also used the existing literature review in trusted repositories to focus the selection of a census on the experience of the information professionals. Background skills and work experience is a determining factor to the success of data collection. The targeted size of the population was informed by the research study. All Information professionals in the NRF organisations were approached for participating in the research study. A total number of 46 Information professionals were approached and only 36 were able to complete the questionnaire.

3.5. Data Collection methods

A questionnaire incorporating closed ended questions was sent out to the participants in this study. The questions were used to collect background information about the participants, whereas open ended questions investigated the experience, skills and training needs of the participants. Based on the existing literature, it was necessary to closed-ended questions to gather both in-depth evidence on the experience of the information professionals and their day to day challenges on managing research data. This structure of questions was necessary to restrict the data collection to specific experiences of information professionals regarding trusted data research repositories. The research selected the use of this instrument to inform successful data collection and analysis.

Zohrabi (2013) discusses close ended, open ended questionnaires, interviews and classroom observations as common methods that can be applied for data collection samples (Zohrabi 2013:254). Johnson and Christensen (2004:312-313) list six major

methods of data collection being tests, questionnaires, interviews, focus groups, observation and constructed and secondary or existing data. Johnson and Christensen (2004) further point out action research reflection as another method of combining all existing major methods in one research study.

3.5.1. Types of Questions

3.5.1.1. Closed-ended

According to Bird (2009: 1311), close ended questions are easy to administer when coding and analysing and these assist with avoiding irrelevant responses from participants. Furthermore, Bird (2009) says that closed ended questions are measured through nominal, ordinal, interval and ratio levels. Wallace and Van Fleet (2012:192) affirm that closed ended questions will limit the respondents to only possible answers.

3.5.1.2. Open ended

Bird (2009:1311) points out some of the advantages of open-ended questions as allowing participants responses to be spontaneous and flexible when answering study questions. Pickard (2013:219) asserts that open-ended questions allow for more detailed responses from participants as the questions are less restrictive.

3.5.1.3. Multi vs single response

Bird (2009: 1311) emphasises that applying both closed ended and open-ended to a research study can yield good results as the research is able to obtain quantifiable and in-depth results from participants. The researcher took this into consideration by applying these types of questions to the study.

3.5.2. Online questionnaires for the study

Wallace and Van Fleet (2012:186) point out the two most popular methods of conducting a research study as mailed and electronic questionnaires. Furthermore, Wallace and Van Fleet (2012) point out that the advantages of using online questionnaires include global reach, flexibility, speed and timeliness and overall ease in the administration of collected data. Online questionnaires also have disadvantages which include low response rate, privacy, inconsistency, lack of clarity and

questionnaires being viewed as spam emails by participants (Wallace and Van Fleet 2012:202).

Bird (2009: 1310) discusses the key features to consider when drafting questionnaires which are questionnaire design, sampling techniques, and data analysis as well as delivery mode. Furthermore Bird (2009) asserts that researchers have to arrange their questions in logical order to allow good transition between questions.

3.5.3. Structure of questions

The researcher is aware of different types of questions that can be applied to a research study. For the purpose of this study, closed ended questions were used. The researcher chose three different themes which contained a set of questions each, and the themes are:

The questions were divided according to the following themes:

Table.2.2: Research themes and questions

Research themes	Research questions
Background work experience	1.1. What is your Business division 1.2. 1.2. What is your position in the Unit/Facility? 1.3. 1.3. Experience in this particular field
Research Data management practice	2.1. What kinds of data do you typically generate in the area of you work? "Tick box" 2.2. Which services do you use for data storage 2.3. Do you create metadata records (Title, abstract, ID etc), manually or automatically?

Research Data storage challenges	<p>2.4. How much storage space do you need for the data you generate?</p> <p>2.5. How long do you require to store the data you generate?</p> <p>2.6. What services do you currently use to store active daily data?</p> <p>2.7. What services do you currently use for long-term data preservation?</p>
Trusted repositories sharing and privacy	<p>2.8. Do you consider the above services to be adequate for your data storage and preservation needs?</p> <p>2.9. Do you use external repositories for data storage? E.g. Google drive</p> <p>2.10. Do you often need to grant access to your data to internal/external people (e.g. collaborators, reviewers, Staff, students) before public release</p> <p>2.11. What kinds of privacy or confidentiality issues do you face? tick as many as apply</p>
Data management training	<p>3.1. Do you require training on data management (storage, access, security)</p> <p>3.2. What area of data management do you need training in?</p>

1. Background work experience - Investigated the background knowledge of the participants in terms of overall work experience.
2. Research Data management practice - Investigated how the participants deal with day to day research data management processes and challenges.
3. Research Data storage challenges - Investigated the existing challenges faced by participants with regards to storing research data.
4. Trusted repositories sharing and privacy – Investigated the level of usage in trusted repositories and privacy concerns
5. Data management training - Investigated the need for participants to be trained in the area of research data management.

The survey consisted of 5 themes with research questions directed at the survey participants. The themes and questions were discussed in detail in this chapter.

Theme1. Background information

The researcher's intention with this theme was to analyse the skills set based on the work experience in terms of the number of years that each participant has. This was to aid the researcher to determine if work experience is a factor when assessing issues relating to research data management. The researcher aimed to investigate if the responses provided by the information professional participating in this research survey was based on individual experience or group experience in the department unit. This question was motivated by the fact that most information professional delegate research data storage to their junior staff members. The researcher formulated the questions below to gather data to respond to this theme:

1. What is your work background?
2. State years of work experience
3. Could you indicate whether you are filling this survey on your own account or on behalf of a group?

Theme 2: Research Data Management

This theme aided the researcher to investigate the available research data in the organisation and the day to day involvement of the Information professionals with generated research data. Furthermore, the background knowledge in terms of structuring and organising research data using metadata was investigated in this theme. Questions in this theme consisted of options that are relevant to the NRF staff to make it easier to select options as they apply to them. The questions below were selected to respond to this theme:

1. What kinds of data do you typically generate in the area of your work?
2. Which services do you use for data storage?
3. Do you create metadata records (Title, abstract, ID etc), manually or automatically?

Theme 3: Research data storage

In theme 3, the researcher asked questions related to the amount of research data generated daily and the storage needs in terms of space for storage. The researcher investigated the level of knowledge related to storage services and preservation. The questions in this theme were extensive as the researcher's main focus of the study relates to the storage of research data. These questions were formulated as follows:

1. How much storage space do you need for the data you generate?
2. How long do you require to store the data you generate?
3. What services do you currently use to store active daily data?
4. What services do you currently use for long-term data preservation?
5. Do you consider the above services to be adequate for your data storage and preservation needs?
6. Do you use external repositories for data storage? E.g. Google drive

Theme 4: Research data sharing and privacy issues

In this theme, the researcher aimed to investigate the application of privacy policies involved when sharing research data. This forms a large component of trusted repositories and forms part of the research questions for the research study.

1. Do you often need to grant access to your data to internal/external people (e.g. collaborators, reviewers, Staff, students) before public release?

2. What kinds of privacy or confidentiality issues do you face? Tick as many as apply?

Theme 5: Research data management training

The researcher used the questions in theme 5 to investigate the training needs of data and information professionals at the NRF. This component was used to determine the gap that exists and advise the organisation in improving the skills of the professionals.

1. Do you require training on data management (storage, access, security)?
2. What area of data management do you need training in?

3.6. Validity and reliability of the instruments

The researcher selected 3 participants from the group who were in senior professional level to assist in the pre-testing of the questionnaire. The researcher used senior level participants to aid in the missing of necessary data to be included in the questionnaire. The participants were also able to assist with simplifying and giving clarity to the overall participants of the study. The aim was to test the questionnaire for errors before it was sent to all the research participants. All necessary changes and amendments were done after the pretesting process. Powell and Connaway (2004:10) emphasizes that pre-testing also allows the researcher to evaluate questions that are not clear to the participants to eliminate any ambiguity.

3.7. Data Analysis procedures

The researcher chose to use Google analytics to analyse the data collected. The data is presented in graphs, tables, numbers and figures to outline all themes identified from the study and to give meaning to the research findings. Furthermore the researcher applied a thematic content analysis approach to the results from the questionnaire. Ngulube, Sibanda and Makoni (2013) assert that with thematic content analysis, research questions form the theme of analysing the data. This is done by identifying recurring patterns from how the research questions are answered. Vaismoradi, Turunen and Bondas (2013) pointed out that the data analysis process is focussed on description and interpretation, modalities of approaches, consideration of

context of data, the data analysis process, and evaluation of the analysis process. Ngulube (2015) points out that data analysis will assist the researcher to design strategies to generate more data to answer research questions.

3.8. Summary

This chapter outlined the methodology approach and research tool used to answer the formulated research questions of this study. Because of geographical reasons, the researcher limited the methodology to a survey that was circulated to the participants in the form of an online questionnaire. The researcher used a detailed questionnaire incorporating closed ended questions with the NRF information and data professionals. The data was analysed by using Google analytics with assistance from a statistician. Copies of all responses from the participants were collected and analysed using the Google analytics software. The responses and analysis of the questionnaire results have been extracted and made available in the investigation. The next chapter will include a data analysis of responses received from study participants.

Chapter 4 Data Analysis, Presentation and interpretation

4.1. Introduction

This chapter focussed on the analysis of the data collected in the study. The researcher collected all responses of the study and analysed them to derive results to the research questions. A detailed overview of charts are presented to present the results of the study. The main purpose of this study was to investigate factors involved in the process of implementing a trusted repository. The results of the survey were analysed to see if they respond to the research study questions.

Hsieh and Shannon (2005) define data analysis as the “process of which sense and meaning are made of the data in qualitative or quantitative research”. Furthermore, data analysis is applied to define and describe data and make comparisons between variables in research to predict outcomes (Hsieh and Shannon 2005). All responses of the survey questions are discussed in detail and are explained in this chapter.

4.2. Background of data collection

The participants for this research study were selected in terms of the NRF departmental units which are directly affected by issues pertaining to research data storage. The units identified were NRF National facilities, RISA, Science engagements, Business and finance systems, HR and Legal unit, SPP, IT Corporate and NRF IT. The researcher based the survey on the day to day work of specialists and the relevance of their work to the research study. 46 Information professionals were identified and selected for sampling and only 36 responded to the questionnaire. The remaining 10 participants could not be reached for the completion of the questionnaire.

Table3.1: Respondents overview

Category	Category size	Category percentage
Selected Population size	46	100%

Respondents	36	83%
-------------	----	-----

4.3. Data analysis and responses

The responses received from 36 participants were analysed using the Google survey analytics also with the involvement of a statistician to help analyse the graphs and tables. The results of the questionnaire were divided in themes and recorded in the same sequence as they were prepared in the All respondent results were presented in a graph or table format where applicable. The researcher provided further analysis of the results to explain the graphs and tables presented in the study.

In the first question all participants were given the consent option to participate in this research study. Loyd (2013: 134) states that in research, allowing consent plays a role in ensuring that participants are aware of their involvement and are happy with the steps taken when gathering data for research.

Q 1.1. What is your Business Division?

Table 4.2: Distribution of respondents by departmental units

NRF Unit	No. Invited to participate	No. of respondents
NRF National facilities	6	5
RISA	14	14
Science engagements	1	0
Business and finance systems	12	11
HR and Legal unit	1	0
SPP	2	0
IT Corporate	4	2
NRF IT	2	1
Knowledge Resources	3	2
Corporate Governance	1	1
Total	46	36
	Percentage (100)	Percentage (83)

Table 4.2 presents the business divisions of the research participants who were invited to be part of the survey and those who were able to respond. 6 librarians from National facilities were selected to participate in the research study, only 5 librarians participated in the survey and 1 was not reachable for completion by the survey closing date. RISA (Research Infrastructure and System Administration) department forms the largest unit in the NRF organisation, 14 information professionals who participated in the research study were from the RISA department. All participants who were targeted from RISA were able to complete the survey. Science engagements, HR & Legal unit, SPP were not reachable to complete the questionnaire. 5 information specialists from IT and Knowledge resources were contacted but only 3 successfully completed the questionnaire. Only 1 participant from corporate governance was consulted and was able to participate in the study.

Q 1.2. What is your position in the Unit/Facility?

Table 4.3 Professional position overview

Professional position	No. Invited to participate	No. Participated
IT Specialist	8	2
Information Specialist	11	11
Liaison Officer	7	3
Librarian	5	5
Professional Officer	8	4
Senior Information Specialist	3	3
Statistician	4	4
Total	46	36

Table 4.3 displays the respondents' professional position within the NRF organisation. This helped the researcher analyse the relationship of participants work experience and their involvement in data management. Information specialists within the NRF are the primary role players when managing research data. There are 8 IT Specialists, 11 information specialists, 7 Liaison, 5 Librarians, 8 Professional officers, 4 Statisticians

and 3 senior Information specialists who were contacted to participate in the study. The researcher selected these role players with the assistance of the NRF Records Manager based on the work experience and NRF research data management.

Q1.3. Experience in this particular field?

Table 4.4: Years of experience

Work experience	Less than 5 years	5-10 Years	More than 10 years
Responses	21	14	1
Percentage	58.3 %	38.9 %	2.8%

In table 4.4, The largest percentage of the information professionals at 58.3% in the research study have less than 5 years of work experience. The majority of the staff members at the NRF is comprised of role players who just entered their professional fields with less than 5 years of work experience in their field of work.

Q1.4. Could you indicate whether you are filling this survey on your own account or on behalf of a group

Figure 4.1: Individual or Group participation

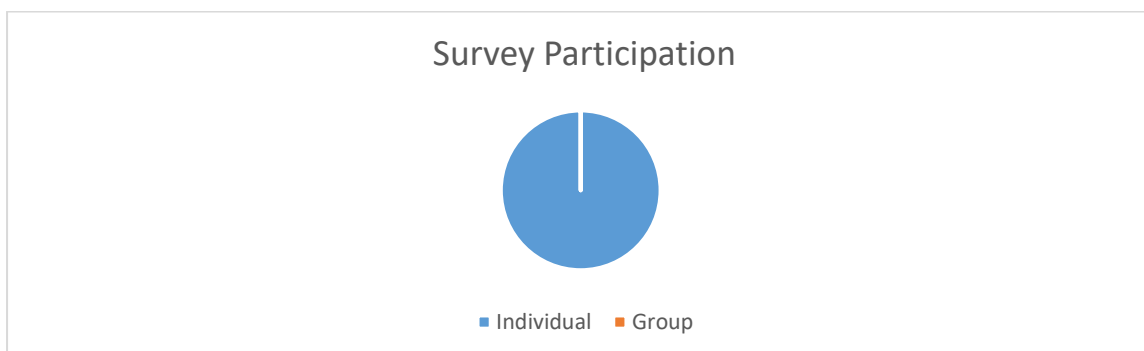


Figure 4.1: All participants filled out this questionnaire on their behalf. The researcher was investigating if the information professionals represent groups at any point while filling out this questionnaire.

Q 2.1. What kinds of data do you typically generate in the area of your work?

Table 4.4: Data types

Data types	Respondents	Percentage
Artefacts	4	12.9
Audio files	2	6.5
Binary raw processed data	5	16.1
Computer code or script	3	9.7
Databases	21	67.7
Gene sequences	3	9.7
Images	7	22.6
Manuals	11	35.5
Movies/Videos	4	12.9
Student data	24	77.4
Reports	21	64.5
Websites	8	25.8
Other files	4	12.9

Table 4.5: The results show that participants generate more student data at 77.4% than all data types. This is followed by databases at 67.7% and reports at 64.5%. The NRF mainly focusses on student funding, therefore the largest kind of data generated in the organisation is student data. Information professionals are also required to make use of databases and reports on a frequent basis to process and store student data. That motivates the high percentage in databases and reports for data that is generated daily. In a similar study, Sewerin et al. (2015) indicated that the most common types of data generated in research projects are numerical data, multimedia and instrument specific data.

Q 2.2. Which services do you use for data storage?

Figure 3.2: Storage services

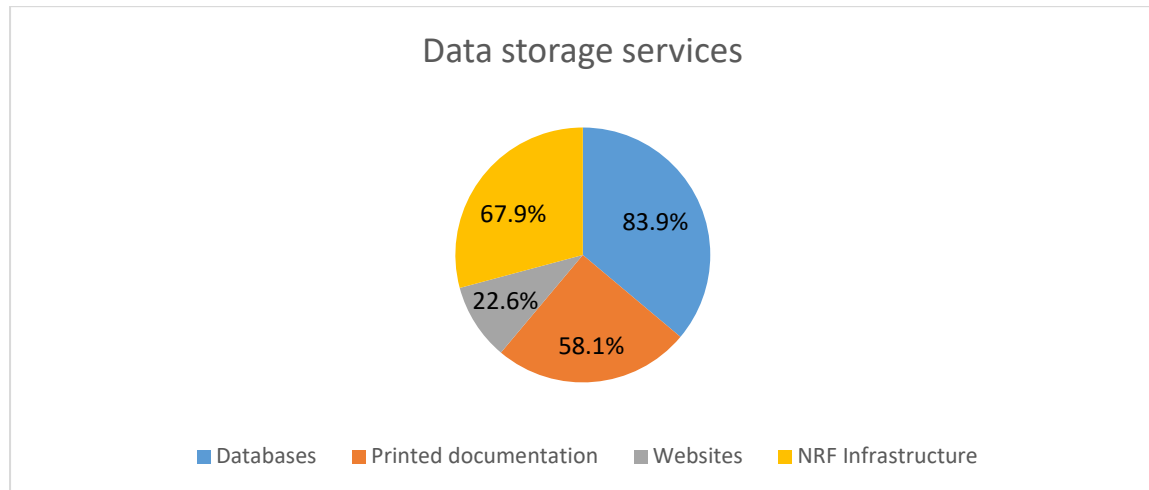


Figure 4.2 displays the responses received in terms of storage services used by the NRF information professionals. With regards to storage, the usage of databases is the most used services at 83.9% and this is motivated by the nature of data that is generated by information professionals in table 4.2.1 which are databases related data.

The NRF infrastructure is the second most used data storage service at 67.7% followed by printed at 58.1%. This is motivated by the NRF records management policy which requires information professionals to store both e-copies on the NRF infrastructure storage and hard copies for filing.

Only a small percentage of professionals at 22% make use of subscription to websites and this might be related to the data confidentiality and sensitivity issues as pointed out by the information professionals in the study. Trust is commonly an issue when storing research data; this motivates the use of internal infrastructure rather than external sources.

Q 2.3. Do you create metadata records (Title, abstract, ID etc.), manually or automatically?

Responses:

➤ ***I do not have any metadata records***

(0%) No information professional selected this option. This shows that all NRF information professionals are aware that metadata is a requirement for all the data they generate and when storing the data. Mowers, Humphrey and Perry (2013) observe that one of the challenges in using metadata is that most professionals are unfamiliar with available metadata standards for research data management.

➤ ***All of my metadata are generated automatically***

30% of the information professionals stated that all their metadata is created automatically. As pointed out in chapter 2 of this study, most storage systems that are created for data storage are metadata ready as they include file names, sizes, locations and restrictions. This allows the information professionals to organize the data accordingly for effective access.

➤ ***Most of my metadata are generated automatically***

30% of the information professionals responded that most of their metadata is created automatically. This shows that to a certain extent the professionals do not follow designed metadata for other types of data generated and stored. Yeumo (2014) points out in a research data management survey that most professionals do not know the benefits of using metadata in managing their research data and furthermore, training and creating awareness is a requirement on the use of metadata in data storage.

About 50% manually / 50% automatically

About 7% of the NRF information professionals stated that 50% of their metadata is automatic when 50% is manually created. This is an indication that there are some information professionals who use the organisation's predetermined metadata as well as their own manually created metadata.

➤ ***Most of my metadata are created manually***

23% of the information professionals create their metadata manually. Most of the information professionals who pointed this out are the same ones that mostly use printed documentation to store their day to day data.

➤ ***All of my metadata are created manually***

10 % of the information professionals stated that all of their metadata are created manually. This can be seen as a challenge as the data can lack standardization in terms of access. Data can also be lost when incorrect metadata is used to store generated data.

Q 2.4. How much storage space do you need for the data you generate?

Responses:

Table 4.5: Storage space requirement

Storage space	Responses	Percentage
	36	100
Less than 10GB	14	38.8
Less than 50 GB	4	11.11
Less than 100GB	6	16.7
Less than 500GB	12	33.3

Table 4.6 shows that most of the NRF professionals (38.8 %) do not need huge storage space for their research data but only need less than 10GB space. The information professionals who selected this option are responsible for mainly generating reports and student related documented data. The other participants selected the need for storage which is between 50GB and 500GB.

Q 2.5. How long do you require to store the data you generate?

Table 4.6: Storage term

Storage term	Responses (36)	Percentage (100)%
Short-term storage (3 Years or less)	15	41.6
Long-term storage (More than 3 years)	21	58.3

More participants pointed out that they will need a period of more than 3 years for the storage of research data. This suggests that preservation strategies will have to be put in place to support this need. This is also motivated by university student data that are normally stored for 2-3 years according to the NRF records policy for student data.

Q 2.6. What services do you currently use to store active daily data?

Table 4.7: Daily data storage services

Responses:	participants 36	Percentage 100%
I use an internal DIY solution, such as an external hard disk/USB	21	58.3
I use the organization provided service, such as the Shared drive	8	22.2
Other/Please specify	7	19.4

Table 4.8 responses display that 58% of the NRF information professionals use an internal DIY solution for long term data preservation. These responses are motivated by lack of trust in the internal organisation provided services. As pointed out by the ANDS (2018), research organisations have to follow the trusted repository elements to ensure security in data storage systems.

Q 2.7. What services do you currently use for long-term data preservation?

Table 4.8: Long-term data storage services

Responses:	Total no:	Percentage
	36	100%
I use an internal DIY solution, such as an external hard disk/USB	6	16.6
I use the organisation -provided service, such as the Shared drive	20	55.6
I use an external service or repository	5	13.9
Other	5	13.9

The information professionals who selected the use of external hard disk were 6 which contributed to 16.6% of the research study. 55.6% of the Information professionals' stated that they use the organisation-provided service, that is, the shared drive. This response is motivated by the NRF records policy which requires most of the day to day generated data to be stored on the organisation's storage systems. Other services used are, external services/repositories at 13.9% and other service preferred personally aside from the listed ones. Averkamp, Gu and Rogers (2014:11) point out the issues that research workers face in terms of storage and opt for services such as external hard drives and 3rd party services.

Q 2.8. Do you consider the above services to be adequate for your data storage and preservation needs?

Table 4.9: Storage and preservation needs

Responses	No of respondents (36)	Percentage 100%
Yes	18	50
No	11	30.6
Other	7	19.4

In table 4.10, 50% of the information professionals responded that the data storage services are adequate for the storage needs. The responses are motivated by the involvement with working with sensitive data that need a secure environment like an organisation internal drive as stated in Q 2.7. Only 30.6% pointed out that the storage service are not adequate for the data they generate and for preservation. This was followed by all other remaining respondents at 19.4%.

Q 2.9. Do you use external repositories for data storage? E.g. Google drive

Table 4.10 External repositories for data storage

Responses	No of respondents	Percentage 100%
------------------	--------------------------	----------------------------

	36	
Yes	20	55.6
No	10	27.8
Other	6	16.7

55.6% NRF information professional make use of external repositories for data storage. The NRF organisation will need to put data strategies in place to ensure data privacy when operating on external repositories.

2.10. Do you often need to grant access to your data to internal/external people (e.g. collaborators, reviewers, Staff, students) before public release?

Table 4.11 Granting access to data

Responses	No of respondents	Percentage
	36	100%
Yes	23	63.8
No	13	36.11

About 63.8 % of the NRF information professionals stated that they are sharing data with external stakeholders on a day to day basis. Griffiths (2009: 51) points out challenges in data sharing such as lack of professional award for data sharing and lack of data sharing resources. Freiman et al. (2010) asserts that other challenges in data sharing include not having funds for data sharing and data sharing requiring too much time to be processed. 36.11% of participants do not share research with external stakeholders.

Q 2.11. What kinds of privacy or confidentiality issues do you face?

Table 4.12: Privacy and confidentiality issues

Privacy issues	Responses	Percentage
	36	100%
Commercial in confidence	14	38.9
Research sensitive	22	61.1
Personal information	17	47.2
None	4	12.9

The findings show that 61.1% of the NRF information professionals are facing data sharing issues related to the sensitivity of research data. This is followed by issues working with personal information at 47.2% as pointed out by the Information professionals. Restriction in sharing sensitive data is motivated by rules and regulations related to working with personal information such as the POPI Act 2013 (Protection of Personal Information Act). Information professionals need to adhere to Acts and regulations at all times when working with personal data. Commercial in confidence at 38.9% is also an issue amongst the Information professionals it can pose intellectual property related issues when sharing organisational information.

Q 3.1. Do you require training on data management?

Table 4..13Data management Training required

Training required	Respondents	Percentage
Yes	36	100
No	0	0

Table 4.14: Theme 5 of the research survey was aimed at addressing the research data training needs of the information professionals. All participants pointed out that

they need training on research data storage, security and cleaning. Research data management is a new discipline in many research organisations. This calls a need for training of staff on research data management. As observed in Patterton, Bothma and Deventer (2018) emerging research workers lack training in research data management. Furthermore the authors point out that training is a crucial part of implementing research data management services in a research organisation.

Q 3.2. What area of data management do you need training in?

Table 4.14: Areas of data management training

Training Areas	Responses	Percentage
	36	100%
Data management training	24	66.6
Data security training	29	80.5
Data cleaning training	21	58.3
None	4	11.11

Table 4.15: Wilson and Patrick (2010: 6) state that “Training should be offered in both ‘broad’ data (or information) management skills and also in ‘narrow’ (technical) data management skills”. Wilson and Patrick (2010) discuss the extensiveness of training that should be offered in research data management; furthermore advisory services is highly recommended in research data management training.

4.5. Summary

This chapter outlined the themes used in the research study. In addition, the chapter also presented an overview of the results obtained from the study survey. The data obtained from 36 participants revealed a gap in the research data management

background and training needs for information professionals. Recommendations from the research results are made in the next chapter.

Chapter 5: Summary of the study

5.1. Introduction

This chapter covered a summary of the research objectives, research questions and findings from the study. A summary of the findings is provided and recommendations made in light of the study findings. The chapter also provides a conclusion of the study.

5.2. Summary of findings and research questions

The researcher applied the identified theoretical framework in the study to guide the research investigation at the NRF. The Data Asset Framework (DAF) stages were applied which are (1) Planning, (2) identifying what data assets exist, (3) information life cycle and (4) providing recommendations for improvement. The researcher used the themes of the research study to present the findings of the survey results. This study aimed at addressing the following research questions:

- i) What is a trusted repository for research data storage?
- ii) What is research data management?
- iii) What data should be stored in a trusted repository at the NRF?
- iv) What are the issues concerning research data storage and preservation at the NRF?
- v) What policies must be implemented to support a trusted repository?

5.2.1. What is a trusted repository for research data storage?

In chapter 2 section 2.2, the researcher outlined requirements that are necessary for organisations to implement trusted repositories. Discussions amongst scholars pointed out these factors that repositories should have:

- *storage and geographic location, file fixity and data integrity, information security, metadata, and file formats.*

- *set of services for storing and making available an institution's intellectual output*
- *responsible for long-term preservation and distribution of research data*
- *authenticity, integrity, confidentiality, and availability of data and services.*
- *complies with integrity, authenticity, legal regulations and quality management*
- *able to identify threats, maintenance and monitoring schedules to ensure successful data preservation*
- *be compliant with the Reference Model for an Open Archival Information System*
- *should also have policies, practices and performance that can be audited to ensure the responsibility it holds of trustworthiness and sustainability*

The NRF information professionals pointed out that they use both internal and external storage services for research data. The internal services used are the Organisations' infrastructure i.e. internal drives and for external services the professionals make use of Websites. This points to the fact that there is a need for a trusted repository for research data storage at the NRF and the above listed factors can be put into effect for successful implementation.

5.2.2. What is research data management?

Chapter 2 section 2.3 discussed in detail a research data management life cycle that summarises all processes related to research data management. An outline of activities that summarises research data management are DataOne (2019):

- *Plan: description of the data that will be compiled, and how the data will be managed and made accessible throughout its lifetime*
- *Collect: observations are made either by hand or with sensors or other instruments and the data are placed into digital form*
- *Assure: the quality of the data are assured through checks and inspections*

- *Describe: data are accurately and thoroughly described using the appropriate metadata standards*
- *Preserve: data are submitted to an appropriate long-term archive (i.e. data centre)*
- *Discover: potentially useful data are located and obtained, along with the relevant information about the data (metadata)*
- *Integrate: data from disparate sources are combined to form one homogeneous set of data that can be readily analysed*
- *Analyse: data are analysed*

Research data management at the NRF

In the questionnaire the NRF Information professionals pointed out that some research data management activities are practiced in the day to day work tasks. However there are no formal procedures in place to guide the application of data storage activities. 10% of the professionals pointed out that they are responsible for assigning metadata to all generated data prior to storage. About 23 % of the professionals pointed out that they are responsible for describing only most of the data they generate when. In terms of data collection the Information professionals collect data as they go without following any formal data collection procedure.

The researcher discovered that there is no research data management framework or data management plan in place to be followed by the information professionals through research data services. With regards to describing data and preserving data, 60% of the information professionals pointed out that either most or all of their research data is automatically described within the organisation. Other processes of the research data life cycle such as data discovery, data integration and data analysing do not form part of the NRF Information professionals' responsibilities.

5.2.3. What data should be stored in a trusted repository at the NRF?

In chapter 2 of the study the researcher presented an overview of literature that supports the intended purpose of a trusted repository in an organisation. All research data that is produced during research activities that is of the benefit of the organisation should be stored in a trusted repository. During the data life cycle the organisation can also make a selection of what data is vital for storage in the trusted repository over a long term or short term. Furthermore, the researcher provided an outline of what a trusted repository should provide in a research organisation which includes the following:

- *storage and preservation of the organisation's institutional memory*
- *research data storage, backup and security*
- *data security and long-term access*
- *making research data accessible to the designated community over the long term*
- *interoperable to existing records management systems in the organisations*
- *ensuring the effectiveness of future life cycle for research data*

Research data produced at the NRF

The production of research data at the NRF is based on the day to day work activities of the information professionals. The Information professionals at the NRF are mainly responsible for processing data related to research grants applications and research programs at the NRF. In the study questionnaire the top 5 types of research data produced by information professionals is (1) student data, (2) reports, (3) manuals, (4) databases and (5) websites. The least produced research data at the NRF are audio files, computer codes, videos/movies and gene sequences. This is due to the type of work done by the information professionals which revolves around processing reports.

5.2.4. What are the issues concerned with research data storage and preservation at the NRF?

The researcher pointed out in the literature that issues experienced in data storage and preservation include the following:

- *finding high quality infrastructure*
- *lack of technologies for implementing a trusted repository*
- *data security and long-term access*
- *regular checks on authenticity and integrity as part of repository administration*
- *trustworthiness, no technical complexity and automatic procedures to reduce human effort*

Research data storage issues at the NRF

The information professionals at the NRF pointed out that they are aware of the research data storage services such as archiving, preservation services and assigning metadata. In the study questionnaire, the researcher provided options for research data storage services which include external hard disk/USB, organisation -provided service such as the Shared drive and external service or repository. 35% of the information professionals pointed out that the storage services are not adequate for their day to day research data storage and do not offer enough storage space. The information professionals also pointed out that there are privacy issues and commercial in confidence issues that need to be taken heed of when storing personal information and sensitive research data. The research data produced has to be properly described and archived to enable access and reuse of the data in future.

5.2.5. What policies must be implemented to support a trusted repository?

In the study, scholars pointed out that the main requirement of policies for trusted repositories is to ensure compliance. The following are the points they enunciated; there is need to:

- *measure compliance of the trusted repository that will house publications*
- *give direction for compliance when archiving digital objects*
- *tool assist in defining metadata, handling data, content, submission and preservation*
- *audit and measure to ensure compliance with relevant standards*

Other studies pointed out that a data management plan is a good starting point towards developing a research data management policy, (NSF 2010) & (ANDS 2018). The informants also pointed out that a research data framework is a good guide to developing a policy (ANDS 2018).

Policies at the NRF

The NRF policies in place used by the NRF information professionals are the records management policy for day to day storage of research related data and the IT policy for the use of systems within the organisation. The policies in place do not include all research data preservation practices and processes which remains an issue for the information professionals. The NRF Open Access statement is another tool used by the Information professionals as point of reference when sharing research data.

5.3. Suggestion for future studies

Future research after the implementation of a trusted repository can ask the following questions:

- *How effective is the repository services to the NRF information professionals?*
- *What are the roles and responsibilities of information professionals in research data management?*
- *What system is suitable for research data management at the NRF?*

5.4. Conclusion

This study was aimed at identifying critical factors for implementing a trusted repository at the NRF. The study reviewed the existing literature to investigate trends, factors and challenges in the area of research data management. The study also formulated research sub-questions to aid in identifying all necessary information to support the main aim of the study. The implementation of a trusted repository at the NRF will require compliance with national research data standards to ensure effective services. This study adopted a quantitative research method to investigate the state of research data at the NRF. A study survey was done with 36 NRF Information professionals to investigate the readiness of the NRF to implement a trusted repository to manage research data. The study revealed some challenges faced by the NRF information

professionals such as storage space, security issues, and training requirements in research data management. Challenges identified should motivate the NRF to consider implementing a trusted repository.

5.5. Recommendations

5.5.1. Implementing a trusted repository

Based on the research data produced at the NRF, it is crucial for the NRF to look into the implementation of a trusted repository for storing and preserving research data. The information professionals also pointed out that the use of external websites for storing research data was useful. The implementation of a trusted repository will ensure that information professionals store, access and share data effectively without issues such as data loss and privacy issues.

5.5.2. Research data preservation

The NRF needs to implement preservation strategies together with the trusted repository to deal with issues of data loss, data security and data sharing. Metadata policy should also be enforced to ensure proper description of research data to be stored in the trusted repository. Metadata will also ensure that data is effectively identified and accessed for future use.

5.5.3. Policy and processes

A research data management plan and research data framework are required to facilitate research data preservation at the NRF. A policy is required for the successful implementation of a trusted repository. The research data management policy and processes have to be integrated into the current existing organisational policies to allow continuation of data services. The NRF as a data generating agency can emulate frameworks such as the OAIS to implement a successful research data framework within the organisation.

5.5.4. Training and awareness

In the study questionnaire all information professionals stated that they require training in the fields of data management, data security and data cleaning. The NRF has to offer training to ensure effective implementation of research data management processes.

5.5.5. Managing public funded research data at NRF

There is a need for the NRF organisation to implement practices and procedures that will guide funded researchers on research data management. A vast amount of NRF research data is generated from publicly funded research which puts the NRF in position to adhere to policies related to the management of research data through trusted repositories. This research study as pointed out literature and empirical work done on issues and challenges faced by research institutions on managing research data. It is vital to support researchers who generate the data and the rights holder through providing a trusted repository. It is therefore recommended that the NRF considers input that was outlined in this study.

References

Akers, K.G and Doty, J. 2013. Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2): 5-26, <https://doi.org/10.2218/ijdc.v8i2.263> (Accessed 15 July 2019).

Alexogiannopoulos, E, McKenney, S and Pickton, M. 2010. Research data management project: a DAF investigation of research data management practices at The University of Northampton. Northampton: The University of Northampton, <http://nectar.northampton.ac.uk/2736/> (Accessed 10 September 2019).

Altman, M. 2006. *An Overview of Digital Libraries: Issues and Trends*. Cambridge: Academic Research Microsoft. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.8357&rep=rep1&type=pdf> (Accessed 12 April 2016).

APARSEN. 2011. Alliance Permanent Access to the Records of Science in Europe Network. www.alliancepermanentaccess.org/index.php/aparsen (Accessed 16 May 2016).

Australian Bureau of Statistics. 2014. Statistical language- Census and sample. {Online}. Available at: <https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Statistical+Language?OpenDocument> (Accessed 25 April 2021)

Australian National Data Service. 2018. Creating a data management framework. https://www.ands.org.au/_data/assets/pdf_file/0005/737276/Creating-a-data-management-framework.pdf (Accessed 08 August 2019).

Australian Research Council. 2013. Research data management. Available at:

<https://www.arc.gov.au/policies-strategies/strategy/research-data-management>

(Accessed April 2020)

Averkamp, S, Gu, X and Rogers, B. 2014. Data management at the University of Iowa: A university libraries report on campus research data needs. A University libraries report on Campus Research Data needs. Available at: DOI: <https://doi.org/10.17077/t5j5-ry3i> (Accessed 07 September 2020)

Ball, A. 2010. Review of the state of the art of the digital curation of research data [online]. Bath: University of Bath. Available at: <http://opus.bath.ac.uk/18774/> (Accessed 17 September 2019).

Ball, A. 2014. 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available at: <http://www.dcc.ac.uk/resources/how-guides> (Accessed online 14 June 2019).

Baškarada, S. 2014. Qualitative case studies guidelines. *The Qualitative Report*, 19(24):1-25.

Bezuidenhout, R and Macanda, M. 2014. *Research Data Management*. Paper presented at the *NeDICC Workgroup*, Pretoria, February. <http://eprints.rclis.org/22717/> (Accessed 20 October 2016).

Bird, D. K. 2009. The use of questionnaires for acquiring information on public perception of natural hazards and risk mitigation - A review of current knowledge and practice. *Natural Hazards and Earth System Sciences*, 9(4), 1307-1325. Available at: <https://doi.org/10.5194/nhess-9-1307-2009> (Accessed 12 August 2020).

Borbasi, S. and Jackson, D. 2012. *Navigating the Maze of Research*. Chatswood, Sydney: Mosby Elsevier.

Boston University Libraries. 2016. Create and manage data: Research data cycle. Available at: <http://www.data-archive.ac.uk/create-manage/life-cycle> (Accessed 16 April 2016).

Budapest Open Access Initiative (BOAI). 2016. *Budapest Open Access Initiative*. Available at: <http://www.budapestopenaccessinitiative.org/read>. Accessed 28 June 2016 (Accessed 12 April 2016).

Burnett, P. 2013. What is the role of a librarian in Research Data Management? Available at: <http://blog.inasp.info/research-data-management-role-librarians/> (Accessed 16 July 2019).

Burns, N, and Grove, S.K. 2011. *Understanding nursing research building an evidence-based practice, 5th ed*, Saunders: Elsevier.

Chervenak, A., Foster, I., Kesselman, C., Salisbury, C. and Tuecke, S. 2000. The data grid: towards an architecture for the distributed management and analysis of large scientific datasets. *Journal of Network and Computer Applications*, 23(3):187–200.

Cibangu, K.S. 2012. Qualitative research: The toolkit of theories in the Social Sciences. In Asunción López-Varela (Ed.). *Theoretical and Methodological Approaches to Social Sciences and Knowledge Management*, pp. 95–126. London: InTech.

Coetzer, G. L. 2018. Designing a geodetic research data management system for the Hartebeeshoek radio astronomy observatory, University of South Africa, Pretoria. Available at: <http://hdl.handle.net/10500/25466> (Accessed 20 April 2020).

Confederation of Open Access Repositories (COAR). 2015. *Promoting open knowledge and open science report of the current state of repositories*. Göttingen: COAR. Available at: <https://www.coar-repositories.org/files/COAR-State-of-Repositories-May-2015-final.pdf> (Accessed online: 12 April 2016).

Consultative Committee for Space Data Systems (CCDS). 2011. *Audit and certification of trustworthy digital repositories*. Washington, DC: Magenta book. Available at: <https://public.ccsds.org/pubs/652x0m1.pdf> (Accessed 12 April 2019).

Coretrustseal. 2019. Data repositories requirements. Available at: <https://www.coretrustseal.org/why-certification/requirements/> (Accessed 12 April 2019)

Corti, M., Van den Eynden, V., Bishop L. and Morgan-Brett, B. 2011. *Managing and sharing data*, UK Data Archive, University of Essex. Available at: <https://ukdataservice.ac.uk/media/622416/trainingresourcespack.pdf> (Accessed 20 June 2019)

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. Available at: [DOI: 10.1177/0961000613492542](https://doi.org/10.1177/0961000613492542) (Accessed 06 August 2019).

Crane, D and Chadwick, I. 2017. Practical strategies for research data management. Available at: http://www.open.ac.uk/blogs/the_orb/wp-content/uploads/2017/10/slides-RDM_PracticalStrategiesForRDM_2017-11.pdf (Accessed 05 February 2019)

Creswell, J.W. 2014. *Research design: Qualitative, quantitative, and mixed methods approaches*. 4th edition. Thousand Oaks, CA: Sage.

Crow R. 2002. *The case for institutional repositories: A SPARC Position Paper*. Washington, DC: The Scholarly Publishing & Academic Resources Coalition.

DataOne. 2019. Data Life cycle. Available at: <https://old.dataone.org/data-life-cycle> (Accessed 05 February 2019)

Deards, K. 2013. Why, how, and where we're going next: A multi-institution look at data management service. *Research Data Management Principles, Practices, and Prospects*. Council on Library and Information Resources, Washington DC.

Dillo, I. and de Leeuw, L. 2018. CoreTrustSeal. *Communications of the Association of Austrian Librarians*, 71(1):162-170. Available at: <https://doi.org/10.31263/voebm.v71i1.1981> (Accessed 20 June 2019).

Digital Curation Centre (DCC). 2011. Trustworthy repositories. Available at: <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories> (Accessed 22 July 2019).

Digital Preservation Coalition. 2019. *Digital preservation*. Available on at: <https://www.dpconline.org/digipres/what-is-digipres>. (Accessed 15 June 2020)

Dobratz, S., Schoger, A. and Strathmann, S. 2007. The nestor catalogue of criteria for trusted digital repository evaluation and certification. *Journal of Digital Information*, 8(2): 2007, Available at: <https://journals.tdl.org/jodi/index.php/jodi/article/view/199/180> (Accessed 06 May 2016).

Donnelly, M. 2014. Data management planning: Overview, policies, resources. In: European Commission data management workshop Brussels, 25 February 2014. Available at: www.dcc.ac.uk (Accessed 02 July 2019)

Downs, R.R. 2017. Improving a trustworthy data repository with ISO 16363. Research Data Alliance (RDA) 10th Plenary. Session: IG RDA/WDS Certification of Digital Repositories Thursday, 21 September 2017. <https://www.rd-alliance.org/sites/default/files/attachment/20170921%20DownsTrustworthyDataRepoISO16363RDAP10Final.pdf> (Accessed 28 August 2019)

Downs, R.R and Chen, R.S. 2013. Independent evaluation of a scientific data center for compliance with the ISO 16363 requirements for audit and certification of trustworthy digital repositories. RDA 2nd Plenary, <http://hdl.handle.net/10022/AC:P:21793> (Accessed 30 July 2019)

Dublin Core Metadata Initiative. 2014. Metadata basics. <https://www.dublincore.org/resources/metadata-basics/> (Accessed 20 July 2019).

Edmonds, W.A. and Kennedy, T.D. 2012. *An applied reference guide to research designs: Quantitative, qualitative, and mixed methods*. Thousand Oaks, CA: Sage.

Faundeen, J. 2017. Developing criteria to establish trusted digital repositories. *Data Science Journal* 16(22): 1-13. Available at: [DOI: http://doi.org/10.5334/dsj-2017-022](http://doi.org/10.5334/dsj-2017-022) (Accessed 18 June 2019).

Flyvbjerg, B. 2006. Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2):219-245.

Fouka, G. and Mantzorou, M. 2011. What are the major ethical Issues in conducting research? Is there a conflict between the research ethics and the nature of nursing? *Health Science Journal*, 5(1):3-14. Available at: <http://www.hsj.gr/medicine/what-are-the-major-ethical-issues-in-conducting-research-is-there-a-conflict-between-the-research-ethics-and-the-nature-of-nursing.pdf> (Accessed 20 June 2019).

Freiman, L., Ward, C., Jones, S., Molloy, L. and Snow, K. 2010. *Incremental scoping study and implementation plan*. Cambridge, Oxford: University of Cambridge, University of Glasgow. Available at: [http://www.lib.cam.ac.uk/preservation/incremental/Incremental Scoping Report 062 010.pdf](http://www.lib.cam.ac.uk/preservation/incremental/Incremental_Scoping_Report_062_010.pdf) (Accessed 30 June 2019).

Giaretta, D. 2009. How does one know which repository is worth its salt? Available at: https://www.researchgate.net/publication/236870426_How_does_one_know_which_repository_is_worth_its_salt (Accessed 25 July 2019).

Gladney, HM. 2012. Long-term digital preservation: a digital humanities topic? *Historical Social Research* 37(3):201-217. Available at: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-378921> (Accessed online 20 October 2016)

Griffiths, A. 2009. The publication of research data: researcher attitudes and behaviour. *The International Journal of Digital Curation*, 1(4): 46-56. Available at: <http://www.ijdc.net/index.php/ijdc/article/view/101/76> (Accessed 22 June 2020).

Halbert, M. 2013. The problematic future of research data management: challenges, opportunities and emerging patterns identified by the DataRes project, *The International journal of digital curation*, 8(2):111–122. Available at: [doi:10.2218/ijdc.v8i2.276](https://doi.org/10.2218/ijdc.v8i2.276) (Accessed 30 June 2019).

Harmsen, H. 2008. The final seal of approval: Directives for data producers/researchers, digital consumers and digital archives. 1st African Digital Management and Curation Conference and Workshop, CSIR Convention Centre, Pretoria, 12-13 February 2008. Available at: <https://nedicc.com/1st-african-conference-for-digital-scholarship-and-curation> (Accessed 21 June 2019).

Hsieh, H.F and Shannon, S.E. 2005. Three approaches to qualitative content analysis, *Qualitative Health Research*, 15(9): 1277- 1288.

Hodson, S. and Molly, L. 2014. Current best practices for research data management policies. A memo for the Danish e-Infrastructure Cooperation and the Danish Digital Library. Available at: <https://zenodo.org/record/27872#.XUgfo1YzaUI> (Accessed 27 June 2019).

Howard, G-M. 2014. 'The literature review'. In: Du Plooy-Cilliers, F, Davis, C and Bezuidenhout, R, eds *Research matters* . Cape Town: Juta, pp. 100–106.

Houghton, B. 2015. Trustworthiness: self-assessment of an institutional repository against ISO 16363-2012. *D-Lib Magazine*, 21(3/4):1-5. Available at: <http://www.dlib.org/dlib/march15/houghton/03houghton.print.html> (Accessed 22 August 2019).

Hugo, W. 2015. Technical requirements data publishing, repository alignment and associated standards. Available at: http://ir.nrf.ac.za/bitstream/handle/10907/131/Wim_technical_2015.pdf?sequence=1, (Accessed 20 July 2019)

International Council for Science-World Data System (ICSU-WDS). 2019. Data and service: Certification. Available at: <https://www.icsu-wds.org/services/certification> (Accessed 25 June 2019).

International Federation of Data Organizations (IFDO). 2016. Data preservation. <http://ifdo.org/wordpress/preservation/> (Accessed 18 October 2016).

Johnson, K. and Christensen, S. 2004. *Educational research: Quantitative, qualitative and mixed approaches*. 2nd edition. Boston: Allyn Bacon.

Johnson, R.B., Onwuegbuzie, A.J and Turner, L.A. 2007. Toward a definition of mixed methods research. *Journal of Mixed Methods Research* 1(112): 112–133. Available at: <https://journals.sagepub.com/doi/abs/10.1177/1558689806298224?journalCode=mmra>. (Accessed 12 July 2019).

Johnson, R.B and Christensen, L.B. 2004. *Educational research: Quantitative, qualitative, and mixed approaches*. Boston, MA: Allyn and Bacon Available at: <http://dx.doi.org/10.3102/0013189X033007014> (Accessed 05 July 2019).

Jones, S. 2011. How to develop a data management and sharing plan. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available at: <http://www.dcc.ac.uk/resources/how-guides> (Accessed 27 June 2019).

Kennan, M.A. and Markauskaite, L. 2015. Research data management practices: a snapshot in time. *International Journal of Digital Curation*, 10(2): 69-95. Available at: DOI:10.2218/ijdc.v10i2.329 (Accessed 20 April 2020).

Koopman, E. and de Jager, K. 2016. Archiving South African digital research data: How ready are we? *South African Journal of Science*, 122(7/8): 1-7. Available at: <https://www.sajs.co.za/article/view/3572/4521> (Accessed 02 July 2019).

Lavoie, B. 2014. The Open Archival Information System (OAIS) reference model: introductory guide. *The Digital Preservation Coalition*. 2nd Edition. Available at: <https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file> (Accessed August 2020)

Lichtmann, M. 2014. *Qualitative research for the Social Science*. London: Sage.

Loyd, D. 2013. Obtaining consent from young people with autism to participate in research. *British Journal of Learning Disabilities*, 41(2), 133-140. Available at: <https://doi.org/10.1111/j.1468-3156.2012.00734.x> (Accessed 25 May 2020)

Lyle, J. 2019. Data repository assessment & certification at ICPSR: Experiences and lessons learned. NIH Trustworthy Data Repositories workshop, Rockville, Maryland, April 8, 2019. Available at: https://datascience.nih.gov/sites/default/files/20190408_NIH_repo_certification_Jared.pdf (Accessed 30 July 2019).

Lynch, C.A. 2003. Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, 226:1-7. Available at: <https://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf> (Accessed 18 October 2016).

Marchionini, G and Moran, B.B. 2012. Informational professionals 2050: Educational possibilities and pathways. North Carolina: School of Information and Library Science of University of North Carolina, Chapel Hill, Available at: <http://sils.unc.edu/sites/default/files/news/Information-Professionals-2050.pdf> (Accessed 2 April 2021)

Mayer, R., Proell, S. and Rauber, A. 2012. On the applicability of workflow management systems for the preservation of business processes, in *Proceedings of the 9th International Conference on Preservation of Digital Objects* (IPRES 2011):101–108, Toronto, Canada, 1-5 October 2012. Available at: https://www.sba-research.org/wp-content/uploads/publications/may_ipres12-workflows.pdf (Accessed 21 June 2019).

Mohajan, H. 2018. Qualitative research methodology in social sciences and related subjects, *Journal of Economic Development, Environment and People*, 7(1): 23-48. Available at: https://mpa.ub.uni-muenchen.de/85654/1/MPRA_paper_85654.pdf , (Accessed 20 August 2019).

Molloy, L. and Snow, K. 2012. The data management skills support initiative: Synthesising postgraduate training in research data management. *The International Journal of Digital Curation*, 7(2):101-109. Available at: <http://www.ijdc.net/index.php/ijdc/article/view/223/292> (Accessed 16 May 2016).

Moore, R. 2008. Towards a theory of digital preservation. *The International Journal of Digital Curation*, 3(1):63-75. Available at: <http://www.ijdc.net/index.php/ijdc/article/view/63> (Accessed 10 May 2016).

Mowers, S, Humphrey, C and Perry CM. 2013. Summary report: Survey of researchers needs and practices regarding research data management in Canada. Available at: <https://gsg.uottawa.ca/data/open/aa-interim-survey-report/20130801-en.pdf> (Accessed 12 August 2019)

National Research Foundation. 2015. NRF open access statement. Available at: <http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded> (Accessed 13 August 2016).

National Science Foundation. 2010. Dissemination and sharing of research results. Available at: <http://www.nsf.gov/bfa/dias/policy/dmp.jsp> (Accessed 10 July 2019).

Ngulube, P. 2015. Qualitative data analysis and interpretation: systematic search for meaning, in *Addressing research challenges: making headway for developing researchers*, edited by ER Mathipa & MT Gumbo. Noordwyk: Mosala-Masedi Publishers & Booksellers.

Ngulube, P. and Ngulube, B. 2015. Mixed methods research in the South African journal of economic and management sciences: an investigation of trends in the literature, *South African Journal of Economic and Management Sciences*, 18(1): 1-13. Available at: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S2222-34362015000100001 (Accessed 21 July 2019).

Ngulube, P., Sibanda, P. and Makoni, NLS. 2013. Mapping access and use of archival materials held at the Bulawayo archives in Zimbabwe. *ESARBICA Journal*, 32: 123-137. Available at: <http://uir.unisa.ac.za/handle/10500/22402> (Accessed 22 July 2019).

O'Dwyer, L.M and Bernauer, J.A. 2014. *Quantitative research for the qualitative researcher*. Los Angeles: Sage

Organisation for Economic Co-operation and Development (OECD). 2015. Making open science a reality. *OECD Science, Technology and Industry Policy Papers*, No. 25, Paris: OECD Publishing. Available at: <http://dx.doi.org/10.1787/5jrs2f963zs1-en> (Accessed 14 July 2019).

Parveen, H. and Showkat, N. 2017. Research ethics. *Communications research*. Available at: https://www.researchgate.net/publication/318912804_Research_Ethics (Accessed 18 June 2019).

Patterton, L, Bothma, T and Deventer, M. 2018. From planning to practice: An action plan for the implementation of research data management services in resource-constrained institutions. *SAJLIS*, 84(2), pp14-26. Available at: <http://www.scielo.org.za/pdf/sajlis/v84n2/03.pdf> (Accessed 09 August 2020)

Pellegrino, J. 2014. A multi-agent based digital preservation model. MA thesis. University of Turin: Italy. Available at: <https://arxiv.org/abs/1408.6126>, (Accessed 12 May 2016).

Peset, F, Fernández, A, Ferrer-Sapena, A, García-García, A, Aleixandre, R, Vidal, A, García, C. 2015. Research data management in Spain: results from a DATASEA project survey. In 10th International Digital Curation Conference, London, 9-12 February 2015. (Unpublished) [Conference poster]

Pickard, A.J. 2013. *Research methods in information*. London: Facet.

Poole, A.H. 2016. The conceptual landscape of digital curation. *Journal of Documentation*, 72(5): 961-986. Available at: <http://dx.doi.org/10.1108/JD-10-2015-0123> (Accessed 02 August 2019).

Powell, R.R. and Connaway, L.S. 2004. *Basic research methods for librarians*. 4th ed. Westport, Connecticut: Libraries Unlimited.

Rankin, J. 2005. *Institutional repositories for the research sector: feasibility study*. National Library of New Zealand Available at: https://docushare.otago.ac.nz/docushare/dsweb/Get/Rendition-68332/IR_report.pdf (Accessed 12 October 2016).

Research Libraries Group- Online Computer Library Center (RLG-OCLC). 2007. *Trustworthy Repositories and Audit certification: Criteria and checklist*. Available at: <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf> (Accessed 22 May 2019).

Resnik, D.B. 2015. *What is ethics in research and why is it Important?* Available at: <https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm> (Accessed 17 June 2019).

Royster, P. 2007. The institutional repository at the University of Nebraska-Lincoln: Its first year of operations. *OCLC Systems and Services: International digital library*

perspectives, 23(2):183-189. Available at: <http://doi:10.1108/1065075071074848> (Accessed 18 May 2016).

Scotland, J. 2012. Exploring the philosophical underpinnings of research: relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, 5(9) Available at: [DOI: 10.5539/elt.v5n9p9](https://doi.org/10.5539/elt.v5n9p9). (Accessed 20 May 2019)

Sewerin, C, Dearborn, D, Henshillwod, A, Spence, M and Zahradnik, T. 2015. Research data management faculty practices: A Canadian perspective. IATUL "Strategic Partnerships for Access and Discovery" Conference, Hannover, Germany, July 5-9, 2015. Available at: <https://tspace.library.utoronto.ca/handle/1807/69145?mode=full> (Accessed 17 September 2020)

Thoegersen, J.L. 2014. Data management services at UNL libraries. Library Conference Presentations and Speeches. Lincoln: University of Nebraska. Available at: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1109&context=library_talks (Accessed 18 May 2016).

Thibodeau, K. 2007. If you build it, will it fly? Criteria for success in a digital repository. *Journal of Digital Information*, 8(2), 2007. Available at: <https://journals.tdl.org/jodi/index.php/jodi/article/view/197/174> (Accessed 10 June 2016).

Thomas, P.Y. 2010. *Towards developing a web-based blended learning environment at the University of Botswana*. D.Ed. thesis, University of South Africa. Available at <http://uir.unisa.ac.za/handle/10500/4245?show=full> (Accessed 10 June 2016).

Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D. and Li, L.J. 2015. *The new data and new challenges in multimedia research*. Available at: <https://arxiv.org/pdf/1503.01817.pdf> (Accessed 02 November 2016).

Thurston, A. 2015. *Managing records and information for transparent and accountable and inclusive governance in the digital era: Lessons from the Nordic countries*. Washington, DC: World Bank. Available at: <http://www.irmt.org/wp2/wp-content/uploads/2015/07/Right-to-Information-Lessons-from-Nordic-Countries.pdf> (Accessed 10 June 2016).

UK Data archive. 2016. *Create and manage data: research data lifecycle*. Available at: <http://www.data-archive.ac.uk/create-manage/life-cycle> (Accessed 12 April 2016).

University of Gloucestershire Research Committee. 2018. *Research ethics: A handbook of principles and procedures*. Available at: <https://www.glos.ac.uk/docs/download/Research/handbook-of-principles-and-procedures.pdf> (Accessed 06 July 2019).

Vaismoradi, M, Turunen H and Bondas T. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3): 398-405. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/nhs.12048> (Accessed 12 June 2019).

Van der Merwe, A, Van Deventer, M and Patteton L. 2015. From process to practice: establishing a research data management in a resources-constrained environment,

Venkatesh, V., Brown, S.A. and Sullivan, Y.W. 2016. Guidelines for conducting mixed-methods research: An Extension and Illustration. *Journal of the Association for*

Information Systems: 17(7): 435-494. Available at: <https://aisel.aisnet.org/jais/vol17/iss7/2> (Accessed 22 July 2019).

Wallace, D.P. and Van Fleet, C. 2012. *Knowledge into action: Research and evaluation in library and information science*. Oxford: Oxford Libraries Unlimited.

Whitmire, A.L., Boock, M. and Sutton, S.C. 2015. Variability in academic research data management practices: implications for data services development from a faculty survey. *Program*, 49(4), 382-407. Available at: doi:10.1108/PROG-02-2015-0017 (Accessed 20 April 2020).

Wilson, A.J. and Patrick, M. 2010. Sudamih Research Requirements Report. Available at: <http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf> (Accessed 20 July 2020)

Wissik, T. and Ďurčo, M. 2015. Research data workflows: from research data lifecycle models to institutional solutions. Linköping Electronic Conference Proceedings, no. 123. Available at: <http://www.ep.liu.se/ecp/123/008/ecp15123008.pdf> (Accessed 30 July 2019).

World Data System Scientific committee. 2015. WDS data sharing principles. Available at: <https://www.icsu-wds.org/organization> (Accessed 27 July 2019).

Yeumo, E.D. 2014. Survey results: Data standards in the Wheat research community. Available at: <https://rd-alliance.org/sites/default/files/RDA%20survey-overview.pdf> (Accessed 22 July 2020)

Zohrabi, M. 2013. Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies*, 3(2):254-262. Available at: <http://www.academypublication.com/issues/past/tpls/vol03/02/06.pdf> (Accessed 18 May 2019).

Appendices

Appendix 1: Research Ethics clearance certificate



DEPARTMENT OF INFORMATION SCIENCE ETHICS REVIEW COMMITTEE

17 December 2019

Dear Ms Mpho Rose Manala

Decision:

**Ethics Approval from 17
December 2019 to 17
December 2024**

DIS Registration #: Rec-051219

References #: 2019-DIS-0051

Name: MR Manala

Student #: 58540008

Researcher(s): Ms Mpho Rose Manala

58540008@mylife.unisa.ac.za

078 974 0188

Supervisor(s): Dr Isabel Schellnack-Kelly

schelis@unisa.ac.za

012 429 6936

**Implementing a trusted repository to support the storage and preservation
of research data at the National Research Foundation (NRF).**

Qualifications: Masters Study



Thank you for the application for research ethics clearance by the Unisa Department of Information Science Research Ethics Committee for the above-mentioned research. Ethics approval is granted for five years.

The **low risk application** was reviewed and expedited by the Department of Information Science Research Ethics Committee on 17 December 2019 in compliance with the Unisa Policy on Research Ethics and the Standards Operating Procedure on Research Ethics Risk Assessment. The proposed research may now commence with the provisions that:

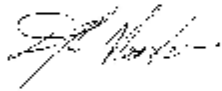
1. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy of Research Ethics.
2. Any adverse circumstances arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the Department of Information Science Ethics Review Committee.
3. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.
4. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards the protection of participants' privacy and the confidentiality of the data should be reported to the Committee in writing, accompanied by a progress report.
5. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no. 4 of 2013; Children's Act no. 38 of 2005 and the National Health Act, no. 61 of 2003.
6. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data requires additional ethics clearance.
7. No field work activities may continue after the expiry date of **17 December 2024**. Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.

Note:

*The reference number **2019-DIS-0051** should be clearly indicated on all forms of communication with the intended research participants, as well as the Committee.*



Yours sincerely



Mr Daniel Rankadi Mosako
Department of Information Science: Ethics Committee



12/5/2019

Data Trusted Repository Survey

Data Trusted Repository Survey

Consent form

I state that I am an employee of National Research Foundation (NRF) and will participate in the Data Audit being conducted as part of the Data Management repository implementation within the organization.

The purpose of the audit is to assess the flow of data in the NRF. A detailed picture of the flow of data in the departments is needed to determine the need for the development of new research data repository and improved systems and processes.

The methodology of the study involves the completion of a questionnaire.

I will be asked to complete a questionnaire, within three days of receiving the questionnaire.

I will also be asked questions about NRF and my experience regarding data Management at NRF.

All information collected in the study is confidential and my name will not be identified at any time.

I understand that I am free to ask questions or to withdraw from participation at any time without penalty.

* Required

1. Do you Consent to participating in this survey *

Mark only one oval.

- Yes
- No

1. Background

2. 1.1. What is your Unit/Facility "choose from the drop down menu"

3. 1.2. How would you describe your role in the Unit?

4. 1.3. Experience in this particular field

Mark only one oval.

- Less than 5 years
- between 5 and 10years
- more than 10 years
- Other: _____

5. 1.4. Could you indicate whether you are filling this survey in on your own account or on behalf of a group?

Mark only one oval.

- Individual
- Group
- Other: _____

12/6/2018

Data Trusted Repository Survey

*Stop filling out this form.***2. Data Management at the NRF**

Encompasses activities that relate to the creation, collection, storage, maintenance, cataloguing, use, dissemination and disposal of institutional data.

6. 2.1. What kinds of data do you typically generate in the area of you work? "Tick box"*Check all that apply.*

- Artifacts
- Audio files
- Binary raw or processed data
- Computer code or scripts
- Computer-readable textual data (e.g. XML files)
- Databases (Access or other)
- Gene sequences
- Human-understandable textual files (e.g. text files)
- Images
- Manuals
- Movies/Videos
- Students data
- Presentations (PowerPoint or other) Spreadsheets (Excel or other)
- Reports
- Web sites
- Other
- Describe
- Other: _____

7. 2.2. Which services do you use for data storage*Check all that apply.*

- Databases
- Printed documentation
- Subscriptions to websites
- NRF Infrastructure
- Other (please elaborate)
- Option 6
- Other: _____

12/1/2018

Data Trusted Repository Survey

8. 2.3. Thinking about your most important work data, to what extent do you create metadata records, manually or automatically?

Mark only one oval.

- I do not have any metadata records
- All of my metadata are generated automatically
- Most of my metadata are generated automatically
- About 50% manually / 50% automatically
- Most of my metadata are created manually
- All of my metadata are created manually
- Other: _____

8. 2.4. How much storage space do you need for the data you generate?

Mark only one oval.

- Less than 10GB
- Less than 50GB
- Less than 100GB
- Less than 500GB
- Other: _____

10. 2.5. How long do you require to store the data you generate?

Mark only one oval.

- Short-term storage (three years or less)
- Long-term storage (more than three years)
- Other: _____

11. 2.6. What services do you currently use to store active in-project (daily) data?

Mark only one oval.

- I use an internal DIY solution, such as an external hard disk
- I use the organization provided service, such as the F: drive
- I use an external service or repository

12. 2.7. What services do you currently use for long-term data preservation?

Mark only one oval.

- I use an internal DIY solution, such as an external hard disk
- I use the organization -provided service, such as the F: drive
- I use an external service or repository
- Other
- Describe
- Other: _____

12/15/2018

Data Trusted Repository Survey

13. 2.8. Do you consider the following services to be adequate for your data storage and preservation needs?

Mark only one oval.

- NRF services for active data storage
- Yes
- No
- Other: _____

14. 2.9. External repositories for data storage?

Check all that apply.

- No
- Yes
- Other
- Please specify
- Other: _____

15. 2.9. Do you often need to grant access to your data to internal/external people (e.g. collaborators, reviewers, staff, students) before public release?

Mark only one oval.

- Yes
- No

16. 2.10. What kinds of privacy or confidentiality issues do you face? Tick as many as apply.

Check all that apply.

- Commercial in confidence
- Research sensitive
- Personal information
- Describe
- Option 5
- Other: _____

3. Summary question

17. Do you require training on data management (storage, access, security)?

Mark only one oval.

- Yes
- No

Powered by
 Google Forms

File: 8/4/2020

MEMORANDUM OF AGREEMENT

Entered into by and between -

The National Research Foundation
(from now on referred to as the NRF)

and

Ms. Mpho Masale
(from now on referred to as the Researcher)

The Parties agree that:

1. The Researcher will treat all information received from the NRF in the strictest of confidence and will not reveal private and confidential information to any third party without the prior written consent of the NRF. The NRF will not permit the Researcher to disclose the names and personal details of the targeted population within Corporate, RISA, and National Facilities.
2. The Researcher will not use the information provided by the NRF for any purpose other than the research project, as defined in the attached correspondence dated 18 May 2020. Any departure from the original proposal must be communicated to the NRF for a re-evaluation of access and approval before implementation.
3. The Researcher will, before the publication of the results of the research, present a final report to the NRF, which is potentially publishable as one or more research outputs for scrutiny to protect confidential information. Approval to publish the theses/dissertation will rest with the NRF and will not be unreasonably refused.
4. The Researcher will submit the final theses/dissertation to the NRF by date of publication for NRF's records.
5. The Researcher will acknowledge the assistance of the NRF in all publications which flow from the information/data.

SIGNED at Pretoria on this 06 day of June 2020

FOR THE RESEARCHER



AS WITNESSES

1. 
2. _____

SIGNED at Pretoria on this 8th day of June 2020

FOR THE NRF



AS WITNESSES

1. 
2. 