DURBAN UNIVERSITY OF TECHNOLOGY
INYUVESI YASETHEKWINI YEZOBUCHWEPHESHE

The Design of a Faculty Research Data Repository Platform conducive to a University of Technology

by

Patiswa Zibani

Student Number: 22173771

A thesis submitted in fulfilment of the requirement for the

Doctor of Philosophy in Library and Information Science

Faculty of Accounting and Informatics, Information Systems, Postgraduate Studies, Durban University of Technology

2023

Supervisor: Dr. F. Marimuthu (Ph.D.)

Co-Supervisor: Dr. M. Rajkoomar (Ph.D.)

DECLARATION

I, *Patiswa Zibani*, declare that:

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

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

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

(iv) This thesis does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

- Their words have been re-written, but the general information attributed to them has been referenced.
- Where their exact words have been used, their writing has been placed inside quotation marks and referenced.

(v) This thesis does not contain text, graphics, or tables copied and pasted from the internet unless specifically acknowledged, and the source being detailed in the thesis and the Reference Section of this thesis.

____

Signature:

Date: 24 May 2023

ACKNOWLEDGEMENTS

First and foremost, I want to express my gratitude to the Almighty God for His favour, grace, mercy, and knowledge bestowed upon me during my PhD programme. I offer Him the first place of praise and glory. Embarking on a journey in pursuit of a doctorate qualification was not an easy one; nonetheless, I value the experience I gathered along the road and would want to express my gratitude to the persons and organizations that, in one way or another, contributed to the successful completion of this research.

My sincere appreciation goes to my supervisors, Dr. F. Marimuthu and Dr. M. Rajkoomar, for their continuous guidance throughout my research project. Your encouragement, support, and positivity have kept me going throughout my study. I would like to extend heartfelt thanks to the Durban University of Technology, Faculty of Accounting & Informatics research participants, postgraduate students, and their supervisors for their insights on faculty research data, which made this study successful.

I am sincerely grateful for the support I received from my colleagues at the Durban University of Technology and in particular to Sara Mitha and the team for the amazing job in formatting the thesis. My appreciation extends to my colleagues at King Abdullah University of Science and Technology (KAUST) who have been helpful through the research project.

To my late husband, Aleck Damson, you left us too early and could not see this project to fruition, but I know you were the Angel that pushed me through the most challenging times; my forever gratitude goes to you. To my kids, Okuhle, Wandile, and Kungawo thank you for your excitement about my study. I persevered not to disappoint you. To my sisters, Siyanda and Pelisa, we have done it Mangwanya Amahle. I would not have completed this project without you. To my brother Aphiwe, thank you for your never-ending support and for standing by my side throughout this journey, mntasekhaya.

Lastly, this study has made me realize that our imagination is the only thing that limits us. Therefore I would like to encourage everyone who wishes to undertake this journey. ***"To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science" -Albert Einstein***

ABSTRACT

As significant players in the production of scholarly work, academic researchers are increasingly gathering and producing data rapidly, exceeding the development of the knowledge and skills required for proper data storage and management. However, the focus has typically been on research publication outputs rather than the research data determining the academic's research output. This study conducted an audit survey on research data management in a University of Technology faculty context to gain insight into their research data management practices. The study was guided by the Data Audit Framework (DAF), the Community Capability Model Framework (CCMF), and the User-Centered Research Data Management Framework (UCRDMF). The study utilized an explanatory sequential mixed-method research design incorporating quantitative and qualitative components. In the quantitative phase, an online survey was administered to postgraduate students pursuing Master's and Doctoral degrees between 2015 and 2020. On the other hand, the qualitative phase involved conducting a meta-analysis of research repositories across global Higher Education Institutions (HEIs) and conducting online interviews with postgraduate supervisors who play a crucial role in postgraduate study and research administration. The quantitative data were analyzed using the Statistical Package for the Social Sciences (SPSS), while the qualitative data were analyzed using NVivo software. The collected data were then analyzed respectively with theoretical frameworks and existing literature. The findings revealed that managing research data was primarily a personal matter. The main reason for the difficulties was that the faculty lacked a research data management investment. The use of emails, external hard drives, and personal laptops are additional examples that showcased that the faculty had not established centralized systems for managing research. However, participants did recognize the benefits of managing research data, such as scientific advances, enhanced data repurpose, and simplicity for data reusers. Findings also showed that the lack of platforms permitting data-sharing and reuse was one of the main reasons that most researchers had not shared or reused other researchers' research data. Researchers, however, wanted to share their data without restrictions. The study's conclusions were also used to create and document a conceptual framework for the faculty research data repository platform. The proposed model aims to guide the conception of a repository that will secure the storage of faculty digital information that can be easily retrieved.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

The most important and frequently used acronyms in this study are listed below:

AIDA	Assessing Institutional Data Assets
CARL	Canadian Association of Research Libraries
CCMF	Community Capability Model Framework
CIF	Crystallographic Information File
CSIR	Council for Scientific and Industrial Research
DAF	Data Audit Framework
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
DUT	Durban University of Technology
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
EU	European Union

FAIR	Findability, Accessibility, Interoperability, and Reusability
FAI	Faculty of Accounting and Informatics
FAQ	Frequently Asked Questions
FITS	Flexible Image Transport System
HATII	Humanities Advanced Technology and Information Institute
HEIs	Higher education institutions
ICT(s)	Information and Communication Technologies
IDMP	Integrated Data Management Planning
JISC	Joint Information Systems Committee
LIS	Library and Information Services
MPEG	Moving Picture Experts Group
NGO(s)	Non-Governmental Organizations
NUST	National University of Science and Technology
OA	Open Access
OAIS	Open Archival Information System

PDF	Portable Document Format
PhD	Doctor of Philosophy
RCUK	Research Council of the United Kingdom
RDM	Research Data Management
RDMS	Research Data Management Services
RTF	Rich Text Format
SA	South Africa
SPSS	Statistical Package for the Social Sciences
SURA	Southeastern Universities Research Association
UCRDMF	User-Centered Research Data Management Framework
UKDA	United Kingdom Data Archive
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UP	University of Pretoria
USGS	U.S. Geological Survey
UoT	Universities of Technology

XML	Extensible Markup Language

CHAPTER ONE: INTRODUCTION TO THE STUDY

1.1 Introduction

As the foundation for creating the faculty research data repository model, this chapter provides a background on research data management (RDM) and describes the study's context. This chapter provides the problem statement, aim and research objectives, breadth and delimitations, and the study's importance. An outline of the study's methodology and research design is also in this chapter. The chapter ends with a description of the thesis' structure.

Scholarly research is essential to national progress since it represents a country's ability to use human resources to solve global challenges (Dora & Kumar 2015: 484). Scholarly research further documents plans to help better understand and address global issues such as health, education, and poverty. Academic research broadens the policy thinking horizon in tackling many of the most pressing issues governments face. The requirement of data is at the heart of most scientific research projects (Dora & Kumar 2015: 484).

According to Ray (2014: 27), research data are essential resources that research institutions must maintain since they are the original sources or material that researchers have developed or gathered when conducting research. Research data can be presented in both digital and non-digital formats. Examples of research data include raw data from a laboratory or a survey; processed data that has been cleaned, improved, structured, and merged in a valuable way; and data published in journals, other related materials, and scientific communication. Dora and Kumar (2015: 484) posit that research data is intricate, comprehensive, unique, expensive, and time-consuming to recreate. Nonetheless, research institutions must be accurate and thorough in collecting, describing, preserving strategies, accessing, re-using, and sharing research data. However, until recently, the focus has been on research-published outputs rather than research data, which defines the outcome of academic research. Different researchers have described the phrase ‘research data’ in various ways. Microarray, numerical and textual records, photos, and sounds are examples of research data that can be used as primary research sources (CARL Data Management Sub-Committee 2009: 4). Borgman (2015: 37) defines research data as information that is gathered, observed or generated for evaluation to generate principal scientific findings. This information can be original data, deduced or evaluated, exploratory or empiric. It can include, but is not restricted to, research lab notebooks, field notebooks, surveys, audio recordings, video recordings, images, samples obtained, specimens, and art forms.

RDM is defined by Ray (2014: 27) as the gathering, categorization, authentication, and storage of data for evaluation, exploration, sharing, re-purposing, and re-invention. Furthermore, Whyte and Tedds (2011: 1) posit that RDM can be viewed as managing data from its introduction into the research process to the publication and preservation of vital findings to guarantee credible results validation and allow revolutionary new research based on current data.

1.2 Research Data Management Perspectives

Across the world, the literature reveals that in recent years, national states, funding organizations, and publishing houses have set the tone and begun implementing stringent open-access data rules and regulations (Corti *et al.* 2014: 60; Nugroho *et al.* 2015: 290). These initiatives began in the United States in 2013 when the White House, through the Office of Science and Technology, authorized ordering all government entities to promote public participation in research and permit open access to data and scientific publications endorsed by public grants. This directive led to the 2018 Foundations for Evidence-Based Policymaking Act, which reinforced the adoption of Open Science principles in the United States (Sheenan 2016: 1; United States Congress 2019).

The European Commission (EU) (2012: 1) published a recommendation on scientific information access and preservation to encourage all its member states to render publicly-funded research findings accessible to the general public to advance science and the knowledge-based economy. According to reports, several research councils in the United Kingdom have Open Access policies dating back to 2005. In April 2013, the Research Council of the United Kingdom (RCUK) mandated that studies supported by RCUK must be published in open-access publications or through self-archiving.

According to the literature (Patterton, Bothma, & van Deventer 2018: 15; Chiware & Becker 2018: 12; Bangani & Moyo 2019: 12), South Africa established the standard for RDM initiatives in Sub-Saharan Africa. The National Research Foundation of South Africa's Open Access statement was a basis for institutions to develop Research Data Repositories (RDRs) as they were mandated to retain research data supporting funded research projects (Patterton *et* *al.* 2018: 23). Furthermore, the South African Department of Science and Innovation released a White Paper that expressed the government's position on research data (South Africa 2019: 19). The paper stipulated government’s plans to promote transparent journal publishing and data exchange; encourage researchers to submit research-related data in publicly accessible repositories; and making publicly funded research data and other scientific reports widely available.

1.3 Overview of Research Data Repositories (RDRs)

Researchers are gathering and producing data at an increasing rate, exceeding the infrastructure development, knowledge, and required skills crucial to enjoying the many benefits of open research data and proper dataset management (Patel 2016: 226). The increased emphasis on research data alters the research culture system. It places new demands on research communities across disciplines (Limani *et al.* 2020: 51). The availability of various research data forms necessitates the development of proper infrastructure to publish and recognize them as research contributions. Research data is becoming more important in the scholarly community, and its development and reuse practices vary amongst research areas.

Various variables influence the evolving research data methods across fields. Several directives from national governments and funding agencies are at the forefront of the push for research data management operations (Critchlow & Van Dam 2016: 223). Due to the increased push towards research transparency and reuse, publishers increasingly require academics to disclose the associated study data for research publications (Krier & Strasser 2014: 3).

Universities and research institutes globally are investing in developing RDRs to increase the availability of information (Shearer, Haigh & Whitehead 2015: 1). RDRs are massive database infrastructures designed to organize, exchange, access, and record researchers' datasets. Repositories might be dedicated and limited to gathering disciplinary or more general data across multiple disciplines, such as the sciences or social sciences (Uzwyshyn 2016: 1). A university or group of universities may provide data to online repositories, which can then be gathered internationally or locally to benefit all parties. The basic concept is that sharing data improves results and accelerates research and discovery. Uzwyshyn (2016: 1) notes that a repository allows other specialists to examine, proofread, review, and validate a researcher's results and the published refereed academic papers. Uzwyshyn (2016: 1) further emphasizes that data research repositories are now required for university campuses and are beneficial for sharing and verifying data-driven research outcomes.

In the Global South, especially in Africa, Open Science (OS), a paradigm intended to make scientific research and data accessible to everybody, has particular difficulties. Although African researchers are willing to use open research techniques, including sharing data, several obstacles have prevented the widespread adoption of OS (Chiware & Skelly 2022:2). These challenges include poor research infrastructures, technological constraints, a lack of financing, policy gaps, expensive open access publishing charges, and a lack of access to high-impact journals. These difficulties considerably influence researchers in the Global South and reflect the broader differences across global societies (Chiware & Skelly 2022:2). Enhancing OS in Africa is a priority in response to these issues. A noteworthy initiative aiming to place African scientists at the forefront of modern, data-intensive science is the African Open Science Platform (Smith & Veldsman 2018: 5). By creating a comprehensive infrastructure that includes hardware, communication, software, policy, and enabling practices, the platform intends to give African researchers access to the tools they need to thrive in contemporary society. The program also created a network of excellence in OS that assisted researchers and other stakeholders in maximizing contemporary data resources for scientific, societal, and financial gains (Smith & Veldsman 2018: 5).

Scholars (Zibani, Rajkoomar & Naicker 2021: 238; Wilkinson *et al.* 2016: 4) have noted that in the long-run, the research environment of today requires that data be findable, accessible, interoperable, and reusable (FAIR). Hence, it is a necessary expectation of the current research environment (universities, funding agencies, publishers). FAIR principles will help Higher Education Institutions (HEIs) develop effective management and administration of valuable digital assets. RDRs should be well-coordinated data management services that keep track of data, starting at the beginning of the research project and continuing through the collection, analysis, documentation, publication, curation, and preservation. Management must fully benefit from all research efforts essential to the optimal repository infrastructure employed by institutions/units (Borgman 2015: 37).

Repositories for research data can be found at Re3data.org., a site that provides a global perspective of RDRs for users with access, publishing houses, library services, and funding organizations. Re3data is a user-friendly platform for scholars to publish and obtain scientific data through their relevant sub-repositories as part of the FAIR Data project and the "CoreTrustSeal" certification. Table 1 summarizes Re3data lists for thirteen countries with 50 or more RDRs. The United States has the most (1144) repositories, followed by Germany (476), the United Kingdom (305), the European Union (285), Canada (262), and France (121). Another 280 RDRs have been designated as International data repositories, and South Africa is far below with only fourteen RDRs recorded.

Table 1.1 Distribution of RDRs by country

Country	**No. of RDRs**
United States	1144
Germany	476
United Kingdom	305
European Union	285
International	280
Canada	262
France	121
Australia	98
Switzerland	85
China	78

Netherlands	69
Japan	66
India	50
South Africa	14

Source: Re3data.org (2022)

Thanks to national governments, funders, and independent organizations, Africa is making considerable progress in championing open data and providing central repositories for access (Olayiwola *et al.* 2016: 101). The successful implementation of open data policies depends on several factors, including the availability of adequate facilities, a significantly higher rate of digital learning, tailored governmental strategies, national leadership, municipal counterparties, local expertise, and a lot of optimism amongst intellectuals, NGOs, ICT businesses and state entities (Olivier 2016: 16). Although institutions have played a significant role in the expansion around scholarly works, there are few RDRs in South Africa as indicated in Table 1. Re3data, a global directory of research data repositories that contain RDRs from many academic disciplines, lists fourteen data repositories from South Africa, including research bodies and organizations. Six South African universities have RDRs with this registry, with only one UoT represented.

Figure 1.1 shows that South Africa is the leading country in RDR adoption, with fourteen repositories accounting for 45%. Kenya, with four, is another country that has achieved tremendous progress (13 %).

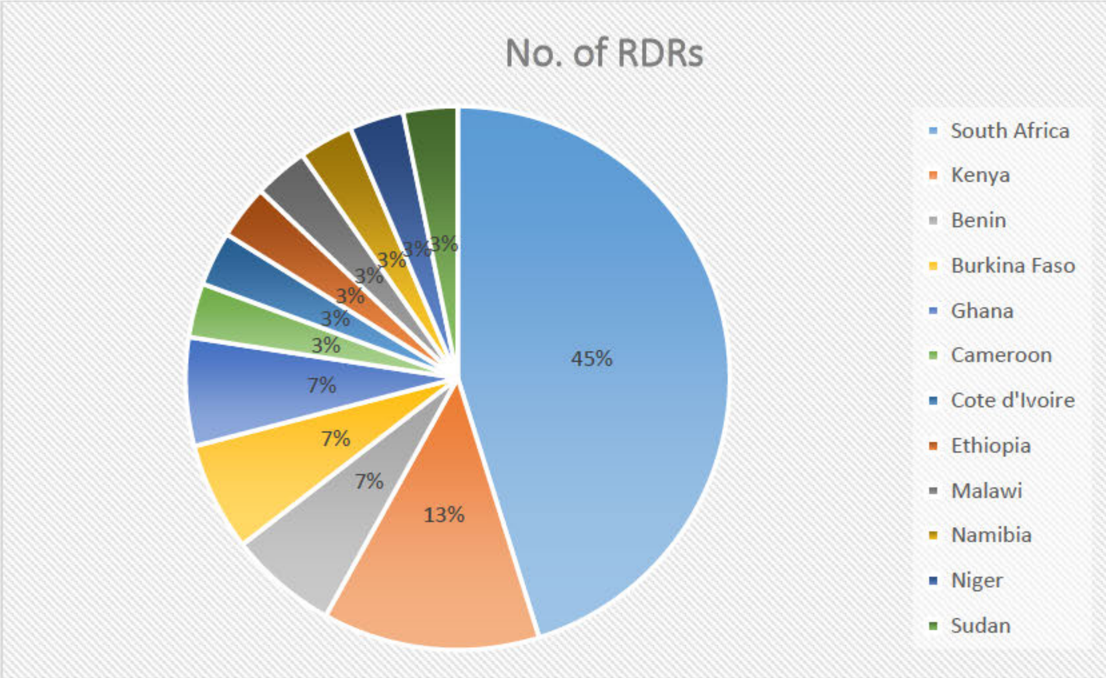

Figure 1.1. RDRs by country in Africa

Source Re3data.org (2022)

1.4 Contextual Setting

South African Universities of Technology (UoTs) arose from the erstwhile Technikons, established to provide career-oriented programmes (Chiware & Becker 2018b: 469). These former Technikons established a solid reputation for delivering career-oriented programmes and served as the foundation for UoTs (South African Council on Higher Education 2010: 6). The only aspect of research in these institutions was of an applied nature, and the goal was to have graduates prepared for the workforce. Due to their hands-on research and connections with business, Technikon courses remained current and relevant, and their alumni were acquainted with how the industry worked through work-integrated learning (SACHE 2010: 6).

South Africa's universities are all on an equal footing in the contemporary higher education environment, but the only difference is their focus. While conventional university research has always been built on social benefit, it has also been grounded on a linear paradigm in which basic research led to innovation, development, production, and societal use. According to Chiware and Becker (2018b: 466), South Africa exhibits a distinctive pattern of knowledge production compared to the rest of the continent. Universities in South Africa have consistently produced more research over the past twenty years, primarily attributable to government incentives for research funding. The Department of Higher Education and Training's (DHET) "accredited lists" prioritize and reward publications in peer-reviewed journals. Furthermore, higher investment in research infrastructures has improved the research output of HEIs. The National Research Foundation (NRF), the main source of research funding in South Africa, which supports postgraduates, postdoctoral researchers, and established academics, is another critical contributor mentioned by Chiware and Becker (2018b: 466).

A UoT emphasizes applied research, innovation, methods, and means of tackling specific challenges in commerce and industry (Chiware & Becker 2018b: 469). No distinction can be made between fundamental and applied research. In many cases, commercial applications enable basic research. The researcher has considered it essential to provide the context for establishing UoTs and what sets them apart from traditional and comprehensive universities in South Africa.

UoTs are fast developing in research, with several appearing in the university rankings that recognize research output and influence worldwide. The DHET-produced Research Output report, released annually, is used to track the growth of research. This report informs the policy for measuring research output at public HEIs. Its goal is to employ incentives to motivate higher education institutions to boost their research production. The incentive program establishes institutional strategic priorities and promotes publishing by giving staff members funding for articles published in academic journals with a review process, making South African academic universities' research available to an international audience (the Republic of South Africa, Department of Higher Education and Training 2022: 11). Chiware and Becker (2018b: 475) emphasize that UoTs contribute to science, technology and innovation in the country's broader educational and socio-economic growth. DUT has established itself as one of the world's most successful research universities. The Republic of South Africa, Department of Higher Education and Training Research Output Report (2022: 18) places DUT as a leading UoT in the following categories: publications output units by publication type; relative cumulative share to sector research publications outputs by individual universities; percentage of book publication output units by a university; growth in weighted per capita research production; and normalized weighted research outputs. As recorded on Re3data.org, only one UoT has an RDR registered. There is enough evidence of a lack of RDM infrastructure within UoTs.

The apparent gap presented within UoTs has prompted the undertaking of this study. The study examines the research data management practices of a faculty in a University of Technology (UoT) context to design a model for a research data repository. The study provides recommendations for creating a prototype for a research data repository platform that may be adopted by other University faculties and used to grow data management principles and practices. The study aspires to support UoTs and all other HEIs in research data management. The study will assist South African universities, and Africa, where open RDRs are still in their infancy and best practices are not established.

1.5 Statement of the Problem

The Durban University of Technology is one of the six universities of technology (UoTs) that emerged from the erstwhile Technikons due to post-independence mergers and re-brandings of South Africa's 36 higher education institutions (Chiware & Becker 2018b: 469). The re-organization of the higher education sector, which began in 2004, saw the number of HEIs reduced to twenty-three (23), with eleven (11) "traditional" universities; six (6) "comprehensive" universities (resulting from fusions involving a technikon and a traditional university), and six (6) universities of technology. Chiware and Becker (2018b: 469) point out that former Technikons were a citadel for career-oriented programmes as their mandate was to educate graduates for the workforce. The UoTs, with diverse advisory boards and Work Integrated Learning (WIL) programmes, were linked to industry to guarantee that their programmes stayed current and relevant and that their graduates were knowledgeable about how industries worked. Research became a part of UoTs programmes later than their counterparts (traditional universities). Therefore, they are considered minnows in the research arena, although some of these universities of technology have made significant progress and are recognized for their global research output and global influence. However, research output is generally not at the output level of traditional universities. According to the Report on The Evaluation of the 2020 University Research Outputs published in March 2022 by the Department of Higher Education and Training, the overall research output for traditional universities is exceptionally higher than that of universities of technologies. Hence, there exists a need for more support mechanisms at universities of technology.

UoTs have benefited from government research funding, created specialized research units, appointed research chairs, and increased postgraduate enrolment and graduates. The postgraduate enterprise is gaining traction at the UoTs, which has enabled the hiring of esteemed researchers and the backing of young scientists' guidance and support (the Republic of South Africa, Department of Higher Education and Training 2016: 39).

Chawinga (2019: 11) concurs with numerous studies that increased research efforts and employing computer programs have resulted in massive amounts of digital research data. Researchers at HEIs publish their findings in both open-access and subscription-based publications. Chiware (2020: 390) highlights the changing landscape of RDM in South Africa, emphasizing the emergence of early adopters within the higher education and research domains. Academic libraries lead in these pioneering initiatives, advocating for and initiating various actions. These include formulating RDM policies and guidelines, establishing essential library and data infrastructures, retraining librarians to deliver RDM services, conducting advocacy and promotional efforts, and aligning with government endeavors to build comprehensive national data-intensive research infrastructures effectively supported by responsive RDM services. In South Africa and Africa, additional advancements in RDM services are primarily focused on specialized subject domain practices that have been established for a considerable duration. These domains operate independently and perceive no need for library involvement, possessing their platforms and frameworks for managing and exchanging data among researchers. This situation is notably observed in domains such as bioinformatics, biomedical research, chemistry, agriculture, and health (Chiware 2020: 390).

An illustrative instance is H3ABioNet, an initiative to enhance African bioinformatics capabilities, specifically facilitating genomic data analysis for H3Africa researchers throughout the continent (Chiware 2020: 390). It was important to note that a significant gap exists as there were no official research data management programmes or services at any faculty in the South African higher education landscape. Numerous studies have indicated that acquiring and retaining data has become an essential component of research data management (Raszewski *et al.* 2021: 160; Gonzalez & Peres-Neto 2015: 436; Vines *et al.* 2014: 95). These studies explicitly argue for the importance of data planning and investment efforts to be recognized throughout and beyond the study initiatives. Research data organized, documented, and stored may be critical to promoting scientific investigation and expanding learning and innovation potential (Knottnerus 2016: 271).

Despite the growing importance of research data and the necessity for proper administration indicated by various studies, institutions have yet to recognize the requirement for systemic data management facilities. For example, Knight (2015: 427) identified a formal research data management infrastructure requirement in his case study. Open Science (OS) in Africa has been the subject of research by academics such as Mwelwa *et al.* (2020: 1), Chiware and Skelly (2022:2), and Abebe *et al.* (2021: 330). They have shown that supporting OS in Africa has the potential to revitalize national science systems and increase their capacities to aid the public and private sectors, as well as the general public. They have, however, noted several obstacles to establishing openness in scientific research. These obstacles include the lack of synergies among African science systems, which frequently function independently. This causes isolated pockets with inconsistent policies, practices, and datasets that lack mutual consistency or interoperability to form (Mwelwa *et al*. 2020: 1). In addition, Abebe *et al.* (2021: 332) assert that "the prospects of open data management and data sharing and its value to the improvement of technological and scientific research in Africa will keep growing, despite the sluggish pace due to inadequate funding, ineffective policy frameworks, and constrained infrastructures." They further clarify how unique the African environment is and how addressing current issues and coming up with solutions will continue to be crucial to Africa's engagement with OS and international open-data projects.

Similarly, Patterton *et al*. (2018: 23) argue that one of the problems with RDM in South African academic libraries and research institutes is the lack of research infrastructure. Patterton *et al*. (2018: 23) also point out that stakeholders in higher learning and research institutions do not always agree on the importance of institutional research infrastructure development. The National Research Fund (NRF) statement was a step towards a national approach to RDM services. Its directive required depositing research data supporting all research outputs funded by them in trusted publicly accessible repositories (Patterton *et al.* 2018: 23). That directive has positively impacted RDM efforts and future policy developments. Whilst universities are growing in research productivity and recognition, there is no evidence of existing RDM practices and frameworks.

1.6 Aim and Objectives

The aim of the study was to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology. The following research objectives [RO] of the study were drawn:

a) [RO 1]: To analyze the most pertinent studies on research repositories at HEIs;
b) [RO 2]: To ascertain the management practices for the existing research data in a University of Technology Faculty;
c) [RO 3]: To establish the usefulness of the faculty research data repositories; and
d) [RO 4]: To design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence.

1.7 Significance of the Study

Increased scholarly research activity has resulted in vast amounts of digital research data that must be stored. However, research datasets may be delicate and vulnerable to storage issues and technological change (Consultative Committee for Space Data Systems 2012: 2). This study is timely because it sheds light on the approaches used by University of Technology faculty members to manage their research data. Universities of Technology in South Africa, by the nature of their existence, never focused on research until post-independence mergers and re-designations of the higher education institutions. According to Re3data.org (2022), the Registry of Research Data Repositories in South Africa shows fourteen registered data repositories, but none belong to a university faculty. This study will close the gap by providing a well-researched design of a faculty research data repository for universities and universities of technology suitable for registration onto the register of research data repositories. Open research data repositories are still growing, with best-practices still to be established. Therefore, much scope exists for this research project. This study's main significance and contribution are that a registered faculty research repository does not exist in South Africa, and the novel design prototype produced by this research study will be of extensive benefit to South African and African universities. It can also be extended to global universities intending to develop a faculty research data repository. The study will also contribute to the body of knowledge to research data management in the field of Library and Information Science.

1.8 Scope and delimitations of the study

The study focused on designing a prototype for a research data repository in a University of Technology Faculty context. A survey tool was used to understand the Faculty of Accounting and Informatics' research data management practices as a case study. The research participants were postgraduate students and supervisors of Master's and Doctoral programmes between 2015-2020. The postgraduate student population was fundamental for their pivotal role in data creation and research data management. The supervisors were selected as key-role players in administering postgraduate studies and research and their capability to provide broader insight into data management practices in the Faculty of Accounting and Informatics. The study excluded undergraduate students and programmes as they are not significant role-players in a study of this nature.

1.9 Outline of Research Methodology

This study investigated research data management practices in a faculty context to support researchers by designing a data repository platform conducive to a University of Technology. The study utilized an explanatory sequential mixed-method research design incorporating quantitative and qualitative components. In the quantitative phase, an online survey was administered to postgraduate students pursuing Master's and Doctoral degrees between 2015 and 2020. On the other hand, the qualitative phase involved conducting a meta-analysis of research repositories across global HEIs and conducting online interviews with postgraduate supervisors who play a crucial role in postgraduate study and research administration. The collected data was then analyzed respectively with theoretical frameworks and existing literature. The overall findings indicated that research data management was predominantly viewed as an individual obligation rather than a structured, collaborative effort within the faculty. A notable challenge highlighted was the insufficient investment in research data management.

The explanatory sequential mixed-method research design used in the study is outlined in Figure 1.2, which provides an adaptation of Creswell and Plano Clark's (2018: 109) mixed-methods design.

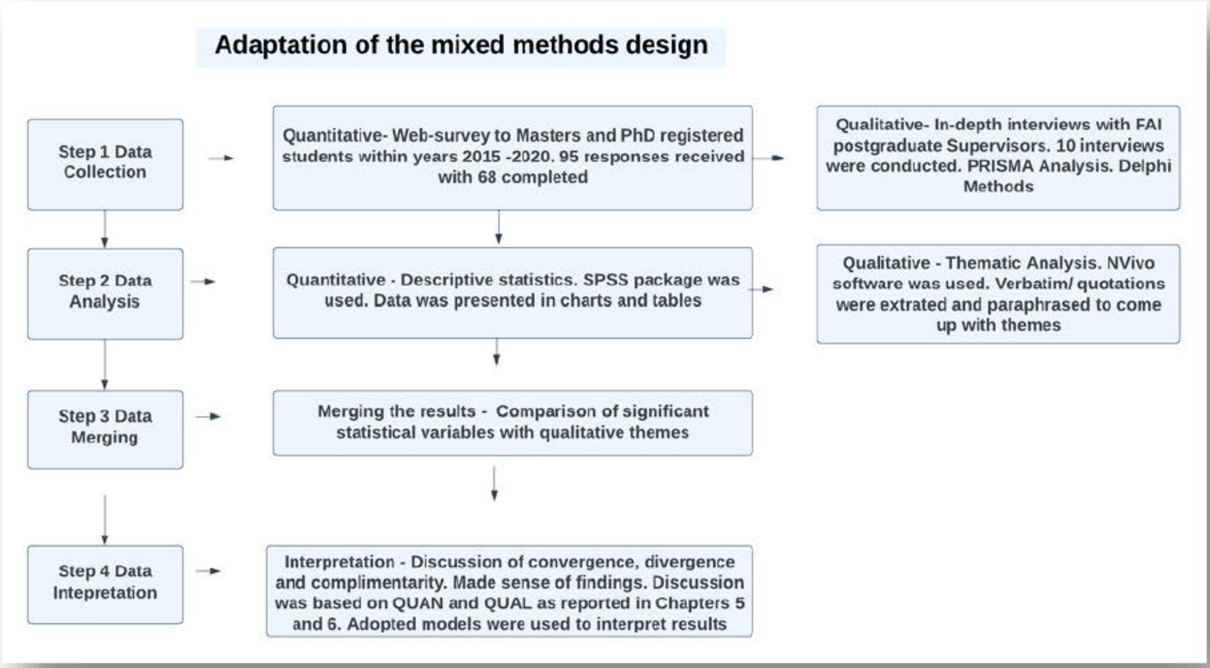

Figure 1.2. Adaptation of the Explanatory sequential mixed methods design

Source : Creswell & Plano Clark (2018:109)

1.10 Research Output

The cited research output below was published in the Emerald journal called Digital Library Perspectives. Zibani, P., Rajkoomar, M. and Naicker, N. (2021), "A systematic review of faculty research repositories at higher education institutions", *Digital Library Perspectives,* Vol. 38, No. 2. https://doi.org/10.1108/DLP-04-2021-0035.

1.11 Structure of the Thesis

This study is presented in seven chapters, arranged in the following manner:

Chapter One: Introduction to the Study

This chapter introduced the research issue by outlining current research data management practices and discussions. The chapter, in particular, provided information on global trends in research data management and its acceptance as a scholarly work attribute. This chapter also defined the research problem, aim and objectives, and the study's significance, influencing the researcher's decision to conduct the study.

Chapter Two: Theoretical Framework

This chapter detailed the three models chosen to guide the current study, namely the Data Audit Framework (DAF), the Community Capability Model Framework (CCMF), and the User-Centered Research Data Management Framework (UCRDMF). The chapter also presented studies that have used the frameworks and clarified the models' selection as the study's foundation guide. There is also a discussion of other relevant models within the research data management discipline.

Chapter Three: Literature Review

The study objectives are the framework for this chapter's comprehensive overview of pertinent print and electronic empirical and theoretical literature. The research-related themes predominant in this field of study are identified in the chapter.

Chapter Four: Research Methodology

This Chapter outlined the study's research setting, methodology, and methods. It specified the study's population, the sampling procedure, the data collection methods, the data analysis, and the presentation procedure.

Chapter Five: Presentation of Quantitative Results and Discussions

The Fifth chapter presents and interprets data acquired through online questionnaires. The study's findings are in narratives, figures, tables, charts, and graphs. This chapter further reviews and analyzes the results reported on the research problem and the study's research objectives.

Chapter Six: Presentation of Qualitative Results and Discussions

The information gathered through interviews is presented and explained in this chapter. The results of the study are presented in narratives and themes. The outcomes related to the research topic and the study's research objectives are further reviewed and analyzed in this chapter.

Chapter Seven: Summary, Conclusions, and Implications of the Study

The final chapter summarized the research, conclusions from the findings, and recommendations. There are also suggestions for further investigation.

1.12 List of Terminology

The following terminologies are defined in the context of this study:

Cluster analysis – Bubble diagrams depict data (keywords) as 'bubbles.' The bigger the bubble, the more frequently words/references are used. Moreover, the proximity of the bubbles indicates that those phrases were related.

Faculty

Entails a collection of university departments with students and academic personnel engaged in teaching, learning, research, and administration.

Hierarchy Charts- Reflect the size of the nodes. The greater the size, the greater the quantity of responses in that area.

Prototype

A pre-production sample, model, or product release designed to validate a concept or method. It is a term applied to various topics, including semantics, design, electronics, and software development.

Research data

Refers to a collection of spreadsheets, papers, photos, videos, and audio supporting the research endeavour, study, or publication.

Research Data Management

Entails active data organization and maintenance throughout the research process and appropriate data preservation at the project's conclusion.

Tree Maps- Display the data (regularly utilized words) regarding block size. As a result, the bigger blocks indicate the most frequently used phrases. The map provides a comprehensive view of how data is positioned to the benchmark size.

Universities of Technology

Refers to the 2004 merger and re-configuration of former Technikons in South Africa to become a part of the higher education landscape. They are popular for producing vocational programmes.

Word clouds- show the most commonly used words. The larger font indicates that, the more frequently the term was used. This aids in identifying vital points.

1.13 Appraisal of the Chapter

In this inaugural chapter, the author adeptly navigated the task of presenting the research's novelty while offering a robust foundation for the study. The chapter's multifaceted objectives introduced the research topic, furnished essential background on methods for managing research data, shed light on data repository platforms, and meticulously defined the context of the study. The achievement of these intentions significantly enhanced the overall value and understanding of the research.

The chapter successfully accentuated the novel aspects of the study, underlining its originality and potential contributions to the field. The research problem was delineated clearly, emphasizing its uniqueness and demonstrating why it warranted dedicated research. This spotlight on novelty was crucial in captivating other researchers' interest and emphasizing the significance of the forthcoming study.

An outstanding feature of this chapter was its ability to summarize the contextualization of the study. The chapter articulated the purpose, objectives, and importance of the study, effectively conveying why the research mattered and the outcomes it intended to achieve. By doing so, the chapter outlined the relevance and timeliness of the study in the broader academic and practical landscape.

CHAPTER TWO: THEORETICAL FRAMEWORK

2.1 Introduction

The theoretical framework is the cornerstone of all research projects' understanding and rationality (Noko & Ngulube 2015: 270). The underpinnings offer a structure or groundwork for the literary works' methodology, analysis, and evaluation and act as the dissertation's investigation's "roadmap" (Noko & Ngulube 2015: 270). The framework provides the structure for defining the philosophy, epistemology, methodology, and analysis strategy for the dissertation and serves as a guide for building and supporting a study. Grant and Osanloo (2014: 12) state that a conceptual model is an architecture that leads research based on a theoretical model using a logical, established justification of occurrences and connections.

Consequently, the preferred school of thought (or theories) is the foundation for the researcher's comprehension of the subject and research planning. Trochim and Donnelly (2006: 18) methodically established criteria for implementing or developing a theory for a doctoral thesis, stipulating that it must be deemed appropriate, comprehensible, and well-connected to the research question. A theoretical framework is a construct created from a model (or models) that already exists in the literature; has been empirically evaluated by others; and has received widespread recognition in the academic literature. A researcher's perspective on the world is what they use (Swanson & Chermack 2013: 75; Rockinson-Szapkiw & Spaulding 2014: 159).

The University of Southern California (2021) concurs with researchers (Trochim & Donnelly 2006: 18; Ravitch & Riggan 2016: 101) that a theoretical framework is not always easily found in the literature. Reviewing research studies to find concepts and analytical models relatable to the currently debated research issue is vital. The theoretical framework strengthens the study in the following ways: When theory presumptions are fully articulated, the researcher can evaluate them critically; the theoretical framework creates a connection between the researcher and prior knowledge and theorizes the theoretical premises of a research investigation, compelling the researcher to respond to why and how queries, moving away from describing the observed phenomenon and generalizing its various elements. A theory helps find the limits to such generalizations since it defines crucial variables that influence an interesting event and emphasizes the need to investigate how and under what circumstances those key variables might differ (Ravitch & Riggan 2016: 101). A theoretical framework, as defined by scholars (Sarter 2006: 493; Maxwell 2012: 96; Grant & Osanloo 2014: 26), is a blueprint that constructs an argument, demonstrates the predicament of the matter, and discusses the findings of a research study.

The models chosen to direct the current study are discussed in this chapter, namely the Data Audit Framework (DAF), the Community Capability Model Framework (CCMF), and the User-Centered Research Data Management Framework (UCRDMF). The chapter also lists studies employing the frameworks and explains why the models were chosen as the study's conceptual basis.

2.2 Foundation models for the study

The aim of the study was to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology. The following research objectives [RO] of the study were drawn:

a) [RO 1]: To analyze the most pertinent studies on research repositories at HEIs;
b) [RO 2]: To ascertain the management practices for the existing research data in a University of Technology Faculty;
c) [RO 3]: To establish the usefulness of the faculty research data repositories; and
d) [RO 4]: To design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence.

The DAF, the CCMF, and the UCRDMF models underpinned the study.

2.3 The Data Audit Framework

The Joint Information Systems Committee (JISC) provided funds for the development of the DAF, which enables all colleges and universities to examine departmental data collections, awareness, policies, and practices for data curation and preservation (Jones *et al*. 2009: 5). It was developed as an online tool and a registry by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow in collaboration with the Digital Curation Centre (DCC). University College London, Imperial College London, University of Edinburgh, and King's College London trialed the model to test and promote its use.

The DAF model is divided into four stages that should be executed in order, as follows:

(a) Planning the audit: identifying and contacting research groups interested in participating in the audit:
(b) Identifying and classifying data assets: learning about the types of data assets created and stored inside an entity;
(c) Examining data asset management: how are data sets stored, maintained, shared, and reused?
(d) Presenting findings: highlight threats such as data loss, misuse, and irreversibility, as well as attitudes toward data creation and sharing and recommendations for improvement and strategies to improve research data management (Jones *et al*. 2009: 5).

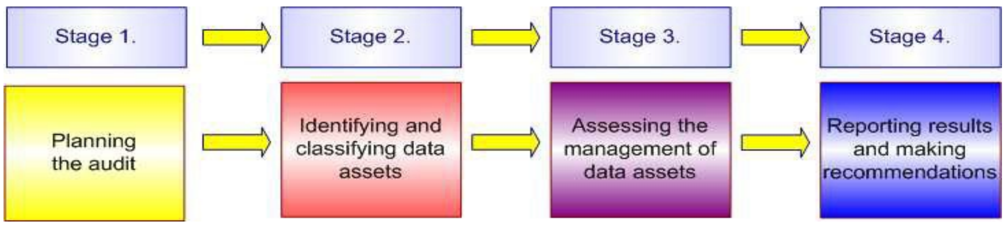

Figure 2.1. The four stages of the Data Audit Framework

Source: Jones *et al*. (2009)

The DAF tool allows institutions to discover their data, where it is stored, and who is (or is not) responsible for it. In essence, DAF enables entities to efficiently collect information about their research data sets. Patterton (2016: 97) agrees with Jones *et al*. (2009: 7) and emphasizes the advantages of using DAF within an organization, namely:

Savings on efficiency

The inability to exchange data holdings information can lead to duplication of work and inefficiency. Organizations can only make educated decisions about how to best use resources for data management if they are well-informed. Auditing data assets provide valuable information that allows proper identification and resource focus. This guarantees that resources are not squandered and not maintaining assets that are not needed. A decision can be easily made to dispose of or relocate such data sets to a more cost-effective offsite storage location. It also informs planning on the appropriate infrastructure and storage (Patterton 2016: 97).

Management of risks

Risks can be managed by organizations that are aware of their data holdings. Poor data management and quality have far-reaching consequences that can be financially, legally, and reputationally devastating. Non-procedural personnel turnover can result in assets becoming orphaned; lowering the value of data assets, their ability to be updated and used, and their long-term survival. Risk management for research data often includes sensitive or personal information, a significant concern. If such data is misused or for which people did not consent, questions about why suitable data management standards were not in place may be raised, and sanctions may be imposed (Jones *et al*. 2009: 8).

Reuse and access

Organizations aware of their research data sets have a better chance of promoting them effectively enough to encourage reuse. This information can be acquired throughout the audit, and it can help consumers find resources more quickly and trust the information they receive. It will also allow for more data exchange if the metadata is interoperable and conforms to regulated vocabularies and taxonomies. Audits assist in determining the worth of assets and identifying underutilized resources (Jones *et al*. 2009: 9).

The framework is designed to be implemented with minimal time and effort, without needing dedicated or specialized personnel. This framework is much more suited to the research since it provides a powerful mechanism for entities to understand their research data sets and where they are to find a suitable and trustworthy repository platform.

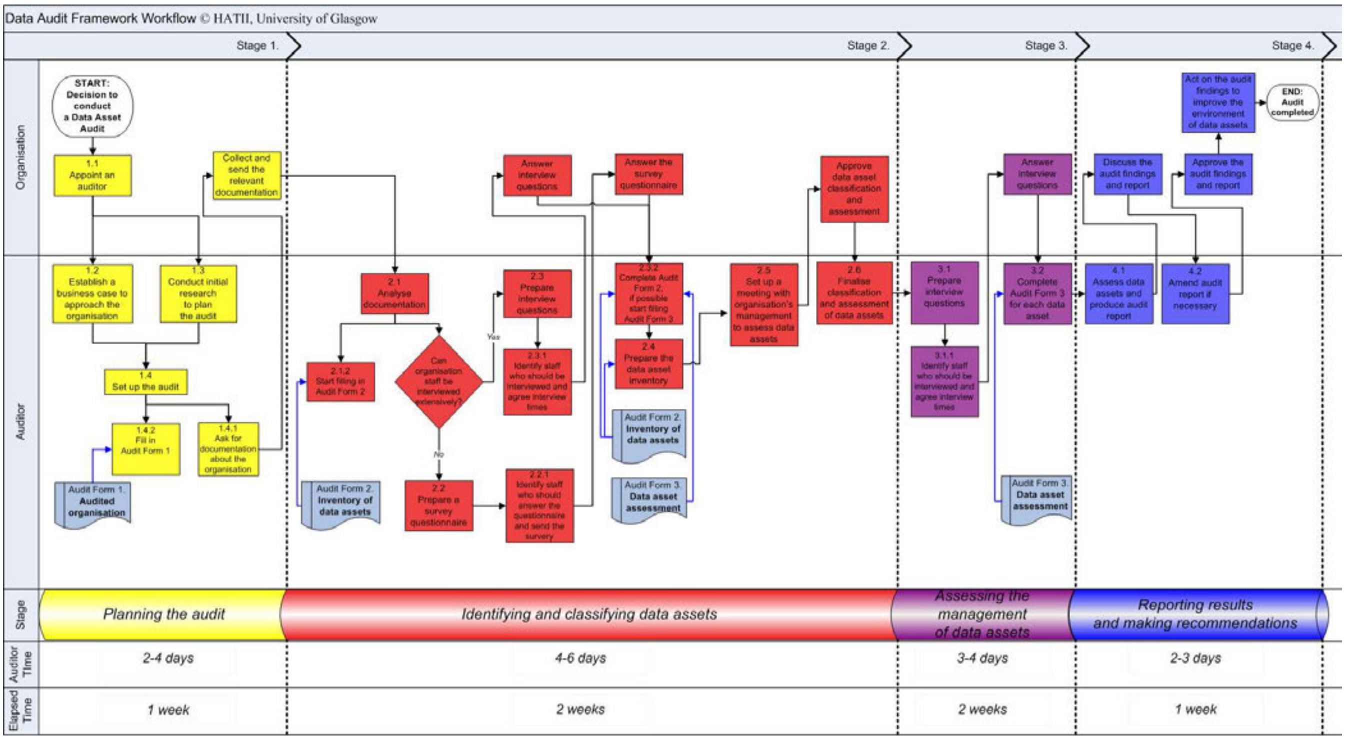

Figure 2.2. The Data Audit Framework Workflows

Source: Jones *et al.* (2009)

2.3.1 Studies that have adopted the Data Audit Framework

Universities worldwide have utilized the DAF to assess research data management methods and readiness.

Oxford University investigated the need for digital repository services to handle and collate the research data created by Oxford scholars using the DAF. The Oxford project's primary objectives covered the following three critical aspects:

A safe, user-friendly method that enables the exchange of massive amounts of data with fine-grained access controls for disciplines that are not currently covered by domain-specific services; a technique to maintain research data over time and a publication platform for fields not yet served by domain-specific services; and advice on how to approach the practical issues to manage scientific data all through the investigation life-cycle (Martines-Uribe 2009: 11).

The DAF methodology assisted in identifying the research groups with which to engage and offered some preliminary data on their data management processes. The framework’s shape and application were changed to match the institution's goals. The DAF served as a helpful planning and execution tool for a strategy to compile data on Oxford research centers' data management activities and assets (Martines-Uribe 2009: 13).

The following were some of the technical considerations gleaned from the Oxford experience:

More guidance on finding research groups that create valuable data should be offered. Perhaps examples of the many forms of data expected to be completed by, for example, clinical trial researchers, crystallographers, social scientists conducting field studies, and so on, could be offered. When presenting findings to research departments, a format for reporting results that ranks research data management practices in a specific department and then provides a visual picture of which data management areas need improvement could be helpful. More information is required to help categorize the data according to its worth, and more knowledge is gleaned on how to disperse value to data (Martines-Uribe 2009: 13).

At the University of Edinburgh, the project was intended to provide Data Audit Framework standards by adapting to the university's data curation needs in research data management methods and inventories (Ekmekcioglu & Rice 2009: 21).

The project's specific objectives were to:

Establish data auditing standards in research centers and schools throughout the three colleges; create and distribute results related to the organizational and technical concerns surrounding the management of research data assets; provide direction, documentation, and training to data creators from various disciplines and across disciplines; make policy recommendations and identify researchers who want to contribute their data through the DSpace repository, web services, or mash-ups (Ekmekcioglu & Rice 2009: 21). The study met its objectives because Edinburgh was one of the institutions that provided standards based on using the DAF. Edinburgh further participated in activities such as writing newsletters and journal articles on lessons learned in the United Kingdom Higher Education. Furthermore, the study was a huge success in Edinburgh's research data management practice and was a critical document that informed RDM strategies and services within the university. On the other hand, it revealed irregular behaviors in many disciplines, particularly in the Humanities, where original data was not created. Researchers relied on published and secondary data sources (Ekmekcioglu & Rice 2009: 21).

After the Bristol Online Survey tool boosted the organization's success in 2012, Knight's (2015: 427) research at the London School of Hygiene and Tropical Medicine used a different approach to data auditing. Similar to the 2012 study, the survey had 15 questions taken from the toolkits for the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) and the data audit framework. While it is acknowledged that these methodologies assist project approaches, the study's questions were re-written to focus on the researchers' standard data management practices for all their ongoing projects. The method successfully gave a broad picture of the ongoing research efforts (Knight 2015: 427). However, inconsistencies were found where respondents had information in many places without stating whether it was the same information duplicated across these locations or information from various research initiatives. The study found that various research stakeholders, such as standards organizations, research funders, journal publishers, federal agencies, and governments, influence data management practices. The study also showed that meeting these requirements required resources and the researcher to re-negotiate agreements because of competing obligations (Knight 2015: 428).

As stated by Alexogiannopoulos, McKenney, and Pickton (2010: 30), the University of Northampton was amongst the initial adopters of the tested data audit framework to investigate the many data kinds that researchers at the institution have, the varied researcher types they use, current data management procedures, and the dangers related to those methodologies. The objective was to gather information in support of a future new data management policy; to establish services based on researchers and funders; and to increase researcher awareness of proper data management procedures (Alexogiannopoulos *et al.* 2010: 30). The survey and interview instruments were utilized in the project. The survey research included various topics, including the types, quantities, and formats of research data maintained; ownership; storage methods; security measures; short-and long-term sharing and access; and funding requirements. The interviews allowed the project team to follow up on significant survey findings and get further technical information on specific data-related queries (Alexogiannopoulos *et al.* 2010: 32). Themes that emerged from this study were: the researchers' distinct needs and characteristics; archiving and preservation requirements; funder mandates; and interest in the university data repository. The study revealed issues and concerns such as data ownership; the use of out-dated data collection formats; data management methods that are more intuitive than ethical; data being neglected once a project is completed; the university's shared server space being underutilized; and researchers being ill-informed, if not misinformed, about the services available to them.

DAF was a successful framework from other parts of the globe and was used extensively. Ndhlovu and Ngwenya (2017: 13) undertook a study to obtain a holistic awareness of the data management procedures of Zimbabwe National University of Science and Technology (NUST) researchers in the Faculty of Applied Sciences and the Faculty of Communication and Information Science. These two academic departments were critical stakeholders in generating, managing, and preserving research data. The study aimed to employ the two faculties to introduce NUST to research data management services. Data for the analysis were gathered for the DAF utilizing questionnaires and in-person interviews. Similarly to the findings of (Alexogiannopoulos *et al.* Pickton 2010: 32), the results revealed the use of obsolete methods of data collection; a preference for sharing data through personal communication over repositories and journals; and concerns about time management when it comes to data preparation (Ndhlovu and Ngwenya 2017: 13).

Similarly, in South Africa, Pienaar (2011: 8) conducted a study at the University of Pretoria (UP) to better understand and enhance research data management methods. Interviews were performed with selected faculty libraries and 52 faculty members. The study revealed five critical areas, namely the funding proposal criteria that varied depending on the budget, where in most cases, no data management or sharing strategies were required; data collection in which UP researchers employed a variety of primary and secondary data collection approaches; and data sets were frequently tiny; and data processing in which researchers at UP primarily used ad hoc data storage on paper and electronically. A few servers were accessible for data storage, but how and where data was saved was primarily up to the individual or department. Moreover, data sharing revealed that UP researchers did not share their data to allow other researchers to utilize it since they did not perceive the need to do so; and support was found to be good across the university (faculty, departments, and research support), but there was a shortage of assistance for data storage, both physical and electronic (Pienaar 2011: 8).

Patterton conducted two further DAF-based investigations two years apart (2014; 2016). These two distinct RDM surveys aimed to determine the existing RDM practices of the Council for

Scientific and Industrial Research (CSIR) workforce. When the initial survey was done, no formalized RDM services, infrastructure, employees, or procedures existed. Personal interviews with thirty-six research group leaders were conducted to determine the RDM behaviors of experienced researchers (Patterton 2016: 68). The second study, which employed an online questionnaire to gather data, was conducted two years later with the participation of 48 young researchers (Patterton 2016: 97). The questionnaire survey was distinct because the polls were just not expected to be a part of the same research and were conducted two years apart. The two surveys revealed behaviors that exposed the institute's RDM procedures and highlighted RDM infrastructure, process, and service shortcomings and limitations. The findings formed the basis for recommending actions to develop the institute's data management service. Overall, and not surprisingly, CSIR researchers found RDM patterns that were out of step with industry standards. Even though research showed that backup frequency was a reliable sign of effective data management, it also showed that DMP deployment had issues with coaching, data maintenance, metadata use, and data storage sites.

Additionally, researchers were requested to highlight any RDM issues or requirements that the survey did not address, and the five significant issues that emerged were: **ICT (Information and Communications Technology).** Although ICT-related difficulties may be considered research infrastructure, several researchers in each group considered storage capacity and data transmission speed relevant; **Data security,** in which researchers expressed concerns about data loss, unintentional data obliteration, data exploitation, encryption flaws, and data loss resulting from equipment theft; **Financial limitations** caused researchers to rely on outdated tools and equipment because software packages, servers and licenses were frequently too expensive for them to afford; concerns were raised about **RDM practices** where researchers struggled to agree on naming conventions, lacked adequate backup knowledge, and had no prior experience in providing metadata. Data integrity, quality control, not using best RDM practices, making data unavailable to research group members, and not using the best RDM procedures were also mentioned; **Data sharing/data confidentiality** concerns regarding the unethical use of data by third parties and the adequacy of the confidentiality safeguards were also significant concerns. Building a mindset shift while dealing with secret or sensitive information is difficult for the newer generation of scientists who are used to releasing all data – open access is the norm for them (Patterton 2018: 23). The study found that individual RDM routines and habits varied greatly, from straightforward RDM practices (naming conventions) to intricate, well-established behaviors like routinely using cloud computing to store well-documented data and adding data to a repository (Patterton 2018: 23).

2.3.2 Key variables in the Data Audit Framework

An entity must take all the necessary steps to evaluate its academic research data collection, awareness, policies, and procedures for data curation and archiving. Three of the actions will be discussed in this section.

Identification, classification, and location of research data assets

The initial phase in the audit procedure is to categorize the department's many types of research data assets and classify them based on their worth to the organization. According to Jones *et al.* (2009: 9), organizations must have correct criteria for identification, classification, and location to fully exploit the potential of research data. An institution/department with suitable mechanisms for identifying, classifying, and locating research data sets is well-positioned to maximize the value of its collections and permit ongoing use (Jones *et al*. 2009: 9).

Research data file formats, size, and management practices

This action examines an entity's or organization's research data file types, sizes, and management procedures to provide a setting for RDM tasks like documenting, archiving, backups, and retention. According to Deloitte (2017: 15), scholars from a wide range of fields are progressively creating digital evidence with intellectual value, such as new types of fellowship, scientific evidence, notations, computer documents, arts, new media, multimedia learning artifacts, user-generated web content, and the outcomes of mass digitization efforts. As a result, it is imperative to create a more thorough inventory list of the numerous electronic research data formats created by researchers from diverse fields (Scott 2014: 21).

Value and maintenance of the research data

This activity comprises determining the worth of research data sets and providing recommendations for managing them based on that value. Scholars (Fecher, Friesike, & Hebing 2015: 18; Guedon 2015: 9; Matlatse 2016: 78; Chen & Wu 2017: 350) assert the notion that due to institutional rules, data inundation, open access proponents, research funders, journal publishers and data enforcement, data sharing has become more and more popular. According to Borgman (2015: 106), most researchers' "basic dilemma" is how to manage their data better. They need "tools, services, and guidance in preserving their data in ways that may be reused, increasing the possibility that their data will be useful to others in the future." According to these authors (Fecher *et al.* 2015: 18; Guedon 2015: 9; Matlatse 2016: 78; Chen & Wu 2017: 350), this activity outlines critical issues that advocate for data openness, such as using research data as a teaching resource, enabling good governance through the formulation of evidence-based policies; minimizing the cost of research data; and serving as the foundation for new research.

2.3.3 Relevance of the Data Audit Framework to the study

The DAF model was chosen for this investigation as the most appropriate model for ascertaining research data management. The DAF was created to let all universities and colleges undertake a data curation and preservation audit of their departments' data collections, awareness, policies, and practices. This feature makes the DAF model a necessary inclusion in this study since it shows how well-informed entities about the research data sets can optimize the worth of their collections by continuing to use them. Furthermore, according to Jones *et al*. (2009: 9), such organizations are in an excellent position to make informed decisions about resource prioritization, which leads to cost savings, managing and realizing the value of data through enhanced access and reuse while minimizing the risks of data loss and irrecoverability. This model was also designed to comprehend existing RDM practices to give mechanisms for managing and sharing data with other entities in a controlled environment, amongst other things. These features set it apart from other traditional research data management methods focusing on curating broad digital objects. This framework is more appropriate for the current study because it will provide a powerful mechanism to understand the understudied faculty's research data and determine a conducive, trusted repository platform.

2.4 The Community Capability Model Framework

The CCMF was created as a self-evaluation instrument to help organizations, funders, and researchers evaluate their capacity to carry out data-intensive research (Lyon *et al*. 2012: 130). This tool is composed of eight critical capability factors that enable the following assessment:

(i) Profiling the community's present state of readiness or capability;
(ii) Identifying critical areas for change and investment; and

(iii) Developing plans for achieving a goal state of readiness.

2.4.1 Studies that have adopted the Community Capability Model Framework

The CCMF model has been used in a sizable number of investigations. The CCMF was created to educate and counsel various research stakeholders on maximizing research output in universities with a focus on research.

Chawinga (2019: 65) states that to create a paradigm for research data management in Malawian institutions; researchers looked into how research data was generated, structured, shared, kept, secured, accessed, and reused in public universities in Malawi. Chawinga (2019: 65) determined that universities were engaged in research activities and that vast amounts of research data were generated due to these studies. The study found that researchers rarely shared their research data because of a lack of infrastructure for data sharing and rewards or incentives. Universities knew the benefits of reusing research data, but doing so was problematic. In general, institutions continued to use sub-par data preservation techniques by backing up and storing data on less dependable free-standing items like personal computers and external hard drives, which were more susceptible to unintentional destruction.

In Ng'eno's study (2018: 49), CCMF was used to assess the research data management practices at a few selected agricultural research organizations in Kenya, intending to recommend solutions to increase agricultural research output management, sharing, and reuse. The model assisted in the understanding of different aspects of research data management at Kenyan research institutes, such as data transparency; skills; and training; technical infrastructure; legal and policy issues; and collaborative research partnerships. According to the study, lacking an RDM legal framework in the institutes assessed resulted in poor research data management. The research data management policies and regulations were outdated, and the functions were not properly coordinated. The study also discovered a lack of research data management understanding and advocacy. Institutes struggled with insufficient standards, guidelines, technical infrastructure, expertise, and collaboration. The study advocated the creation of a formal data governance organization to handle research data management challenges; a framework for legislation and policies; strategies for creating capability; and a top-notch technical infrastructure to encourage research.

Shen (2015: 172) used the CCMF to investigate Virginia Tech researchers' data-sharing practices. According to the report, researchers did not fully embrace open data-sharing due to various circumstances, including a lack of funds, restricted time, and a lack of incentives. The study discovered a considerable gap between data-sharing activities and significant and beneficial reuse or re-purposed merits, implying that data's real potential for future research was lost immediately after the original study was completed. Researchers' erratic and limited data management and documentation methods also contributed to their challenges when re-using existing data.

Lyon, Patel, and Takeda (2014: 4) used the CCMF to determine the prerequisites for RDM in academic libraries by collecting data from e-Science researchers from Cambridge (UK), Melbourne (Australia), Stockholm (Sweden), Bristol, York (UK), and Amsterdam who engaged in an international workshop (Netherlands). This was done to comprehend the researcher's various maturity patterns and scope in the research data requirements environment. Based on the research findings, the researchers created summaries and infographics of data-intensive potential. This led to the creation of a new multi-faceted capability instrument for designing and managing customized RDM programs in university libraries.

2.4.2 Key variables in the Community Capability Model Framework

The CCMF model identifies eight critical factors: collaboration, skills and training, openness, technical infrastructure, standard practices, economic and business models, and legal and ethical issues. These factors are crucial for evaluating an organization's, funder's, or researcher's capacity to conduct data-intensive research (Lyon *et al*. 2012: 128).

Collaboration

Lyon *et al*. (2012: 127) explained that the collaborative context exposes working relationships developed based on research. These collaborations vary in kind and size, some are more formalized and governed by contracts and agreements, whilst others are more informal and self-managed (Lyon *et al*. 2012: 127). This activity was not utilized in the study. Therefore, this variable was not employed in the current investigation.

Skills and training

According to Lyon *et al.* (2012: 127), the community's ability to perform rigorous research is greatly determined by its members' particular skills. As a result, community competence can be enhanced by providing members with relevant skills training. "Tools and technologies such as cloud technology, visualizations, statistical methods, simulations, and modeling; data synopsis and identification such as metadata, vocabularies, and corroboration; and policy and planning such as data management plans, policy frameworks" are some examples of such skills sets (Lyon *et al*. 2012: 128). This variable was used as it linked to an understanding of research data types, classification, special features, storage, etc.

Openness

It is vital to have open access to research data to advance science, scholarship, and society. The open-access movement, through declarations such as the Berlin Declaration (2003); European Commission (2012: 3); UNESCO Open Science Movement (2021: 11); and the South African White Paper on Science, Technology and Innovation (2019: 19), have advocated for research data to be curated and made widely accessible through open access to improve research data use and reuse. According to Lyon *et al*. (2012: 128), openness is the free exchange of methodological approaches and findings, which advances knowledge. The idea of openness is used throughout the research process, especially when conveying study plans, summarizing ongoing research, and making available existing material to a broader audience (Lyon *et al*. 2012: 128). The research community is progressively making requests for researchers to open up the research data workflow process from inception to methodology and final results and conclusions to bring the addition of significance to the whole process of research through the verification, replication, and reuse of findings (Lyon *et al*. 2012: 129). This variable is used in the study as it addresses the ultimate goal of identifying research data sets to facilitate their openness.

Technical Infrastructure

Every RDM endeavor needs a solid infrastructure to initiate with success. Shakeri (2013: 15) states that data curation without the mandatory infrastructure is impossible. At all stages of the research life-cycle, the technology infrastructure makes research instruments and services more convenient (Lyon *et al*. 2012: 129). Connectivity, engagement, citizen science, data discovery, access systems, documentation and handling, storage, archiving and conservation, simulation, and analysis exist. (Lyon *et al*. 2012: 129). For data research administration, computer-based storage facilities are required. This activity was covered in the UCRDMF model utilized in the study, and this variable was not employed in the current investigation.

Common practices

Lyon *et al*. (2012: 129) assert that research teams have established unofficial guidelines in certain aspects, including data formats, data gathering methods, computation processes, data labeling, exchange guidelines, statistical modeling, concepts, schemas, vocabularies, and data signifiers. Data standards should be valued for quality, especially for promoting and facilitating data reuse and combination. Within specific research fields, researchers must share and understand these standard criteria (Lyon *et al*. 2012: 130). Therefore, this variable delved deeply into common standards used by researchers within RDM frameworks and was covered in the study.

Economic and business models

Lyon *et al*. (2012: 130) argue that data-driven science necessitates financial investment and that looking at long-term funding options is crucial. Significant financial investments in infrastructure and time series data reviews in the social sciences are primary investment areas and significant central investments in network infrastructure (Lyon *et al*. 2012: 130). This variable was not used as it relates to large-scale investments that resided outside the scope of a faculty/ unit and, therefore, was not found relevant to the study.

Legal and ethical issues

Lyon *et al*. (2012: 131) aver that the legal and ethical aspects of legal, policy, and regulatory frameworks and ethical considerations are crucial. In research, ethical responsibilities may prevent particular datasets from being shared. Legal issues may prevent data from being shared in the first place, let alone re-purposed or reused. As an illustration, informed consent signed by patient populations severely restricts what might be done with the output data in clinical research (Lyon *et al*. 2012: 131). Gray, Grove, and Sutherland (2016: 173) stress the significance of anonymity and voluntary participation in most individual-related data. Legal and policy frameworks for RDM are necessary to direct the creation, evaluation, categorization, preservation, sharing, and reuse of the intellectual property and ethical considerations of research data (Lyon *et al*. 2012: 131). This variable was used as it related to research data access, sharing, reuse classification, and storage issues.

Academic culture

According to Lyon *et al*. (2012: 131), significant research is more probable to be effective in settings where data are greatly valued, scientists are recognized for their significant contribution, and strict criteria are anticipated for data entering the research log. The CCMF demonstrates the success of reward models for researchers in which all efforts are acknowledged and respected per set norms and indicators. This opens the door to other avenues of welcome and support, such as entrepreneurship and innovation (Lyon *et al*. 2012: 131). This variable was not used for this study.

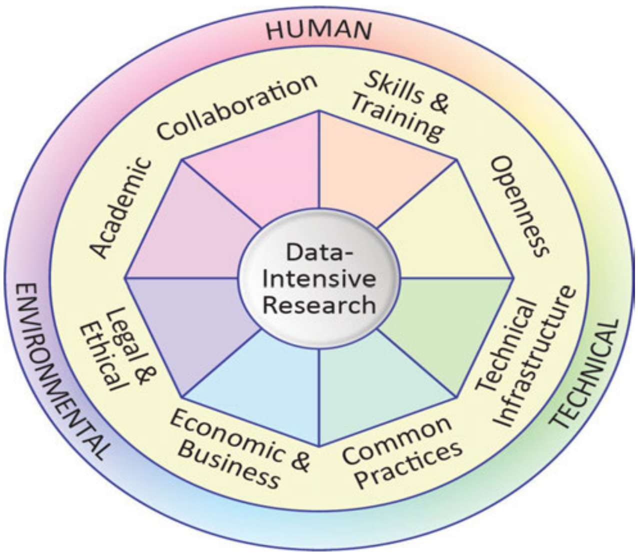

Figure 2.3. The Community Capability Model Framework

Source: Lyon et al. (2012)

2.4.3 Relevance of the Community Capability Model Framework in this study

As indicated at the beginning of this section, the CCMF has been created as a self-evaluation instrument for institutions, funders, and researchers to assess their ability to conduct data-extensive research (Lyon *et al*. 2012: 126). The CCM framework specifies the roles, responsibilities, and requirements for each capacity to improve the effectiveness and efficiency of RDM. This is especially important in faculty-generated research data, as data curation concerns digital research data.

2.5 The User-Centered Research Data Management Framework

This new research framework presents a comprehensive approach to designing a research data management system that considers the system users' prerequisites and the particular specifications for research data. The framework illustrates the creation of a prototype of a user-centered, data-conscious research data management system (Bugaje 2019: 13). User-centered design is an essential aspect of research data management, so it should be considered in developing all research data management systems and connected data. As attested by Bugaje (2019: 13), in research and scholarship, data and publications go together, and a cursory examination will show that everything in the research data management ecosystem—from data sharing on repositories to identifying and perhaps re-using it—is based upon or connected to the user or the user's authorization. This model is less than two years old; thus, other researchers are unlikely to have extensively used, tested, and rigorously validated it. However, the new model does not preclude other researchers from using it. The User-Centered Research Data Management Framework is divided into three phases:

a) Phase I- information gathering to comprehend and specify the use context; specify the needs of the user;
b) Phase II- prototype design and development to provide design solutions; and
c) Phase III- system evaluation to evaluate the strategies against the requirements

d) Phases 2 and 3 from this framework are suitable for the research as they will guide the design of a model for the faculty repository context.

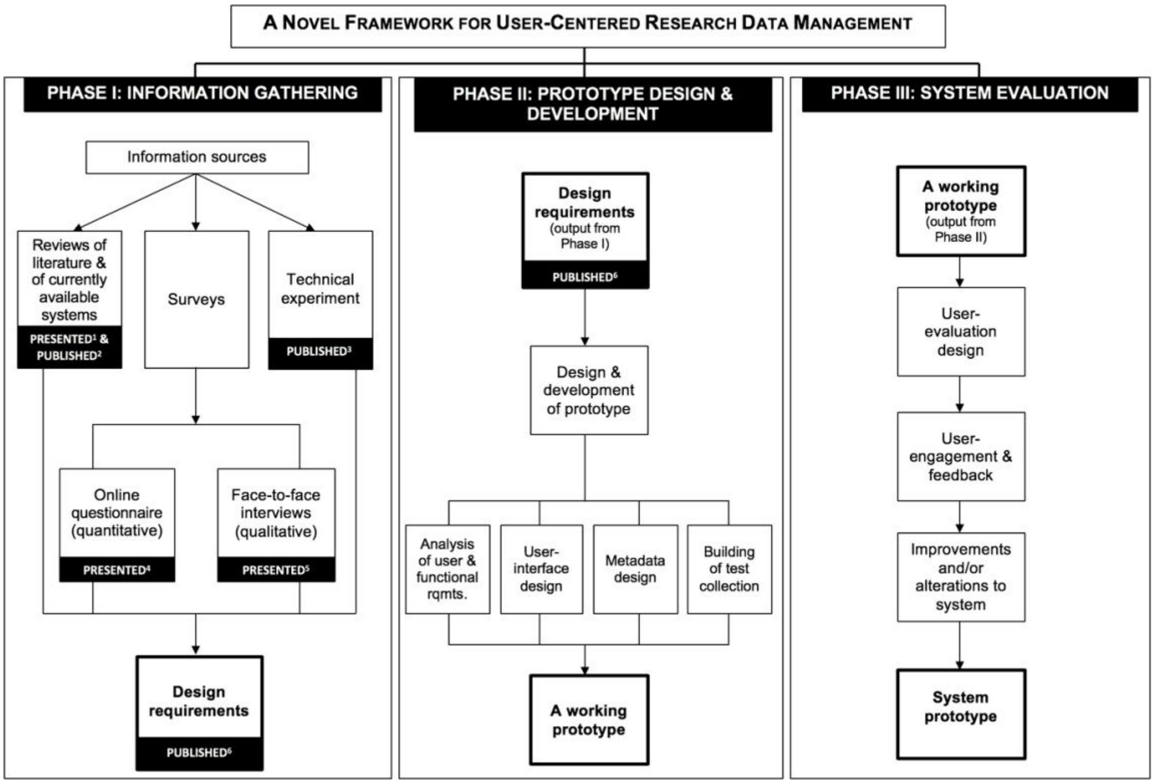

Figure 2.4. The User-Centered Research Data Management Framework workflow

Source: Bugaje (2019)

2.5.1 Studies that have adopted the User-Centered Research Data Management Framework

As previously stated, the User-Centered Research Data Management Framework is a newly established model that aims to design a research data management system informed by the demands of system users and unique research data requirements. Bugaje (2019: 13) used this paradigm in a study for the University of Northumbria in Newcastle to design a modest model of a user-focused and data-conscious research data management system called DataFinder. Since this model is only a few years old, it is unlikely that it has been generally accepted, tested, and thoroughly confirmed by other academics. The study showed two essential issues: the relevance of a system that addresses user-specific needs and linked data (Bugaje 2019: 148). The study advocated developing a research data management system incorporating user-focused enterprise and related data in a relatively simple execution.

2.5.2 Key variables in the User-Centered Research Data Management Framework

The three core actions for a simple user-centered research data management system are unpacked in order and in line with the base activities common to user-centered design process models (Zimmermann & Grötzbach 2007: 98; Bugaje 2019: 13).

Information gathering

The user's information collection techniques and system requirements for research data management are covered in this activity. These techniques include content analysis, surveys, and interviews (Bugaje 2019: 102). However, because this activity was covered in the DAF utilized in the study, this variable was not employed in the current investigation.

Prototype design and development

The functional (user), technical nature, additional requirements, and considerations linked to developing research data management systems, such as metadata and persistent identification, are all discussed in depth in this activity. A collective analysis of user requirements is essential in the three main phases of user-focused process development (Ames 2001: 10; Lazar 2006: 98; Satzinger, Burd, & Jackson 2016: 4) because it outlines the objectives of a new system and provides a specification for achieving them. Any RDM project needs a robust infrastructure to get off the ground and succeed. Shakeri (2013: 15) states that data curation cannot be done without the required infrastructure.

System evaluation

Zimmermann and Grötzbach (2007: 98) noted that system evaluation validates user needs and improves system design. It aids in testing basic concepts and identifying the system's strengths and weaknesses by actions aimed at collecting arbitrary user experiences. System evaluation activity is discussed under this variable.

2.5.3 Relevance of the User-Centered Research Data Management Framework in this study

A large-scale need for research data repositories has resulted from the global drive to adopt and advocate Open Data principles. RDRs are portrayed as a critical tool and solution for research data management, collaboration, and distribution, assisting in realizing the benefits that drove the policies. In most circumstances, existing RDM technologies and infrastructure are insufficient to support and advance this objective. This framework is essential because it presents a holistic paradigm for RDM system design that explicitly considers research data requirements and system users' specific needs to develop effective systems. Through the variables it presents, such as information collecting, creating a prototype, and assessing the system, the UCRDMF shows the creation of a straightforward prototype of a user-focused, data-conscious research data management system, which is found to be helpful for the study.

2.6 Mapping the study objectives to the Selected Models

The table below summarizes how the attributes gathered from the three adopted models are tested in this research.

Table 2.1. Mapping the research objectives to the adopted models

Study objectives	**Attributes covered by the study**	**Sources of attributes**	**Reviewed studies that used the DAF, CCMF, and UCRDMF models**
To analyze studies on research repositories at HEIs using machine learning	Research repository types, classification, storage capabilities, special data features, open access,	DAF model, CCMF model, literature review	• Oxford University Libraries (2009) used the DAF model to assess research data management methods and readiness. • Ng'eno (2018) used the CCMF model to examine the research data management procedures at some agricultural research institutes in Kenya.

To ascertain the management of the existing research data	Data formats, data storage facilities, data management, metadata, preservation, standards, migration, preservation activities, servers,	DAF model, CCMF model, literature review	• Ndhlovu and Ngwenya (2017)used the DAF model to get a holistic awareness of the data management procedures of Zimbabwe National University of Science and Technology (NUST) researchers in the Faculty of Applied Sciences and the Faculty of Communication and Information Science.
To establish the usefulness of faculty research data repositories	Research data use, sharing data, policies, regulations, openness,	DAF model, CCMF model, literature review	• Patterton (2016) used the DAF model to determine the existing RDM practices of the Council for Scientific and Industrial Research (CSIR) workforce. • Lyon, Patel, and Takeda (2014) gathered information from e-science researchers who took part in a global conference from Cambridge (UK), Melbourne (Australia), Stockholm (Sweden), Bristol, and York (UK), as well as Amsterdam, using the CCMF to ascertain the preconditions for RDM in university libraries (Netherlands).

To design a novel digital prototype for a faculty research data repository platform based on research evidence	Prototype design, user-interface design, prototype development, prototype presentation, prototype review	UCRDMF model, CCMF model, literature review	• Bogaje (2019) used the UCRDMF model to propose an all-inclusive model for RDM system design that considers system users' needs and specific research data requirements. • Lyon, Patel, and Takeda (2014) investigated requirements for research data management support in university libraries using the CCMF model, presenting a novel multi-faceted capability tool.

2.7 Justification for the models adopted in the study

This study used the Data Audit Framework, the Community Capability Model Framework, and the User-Centered Research Data Management Framework. The three models were adopted because they complemented and strengthened one another in various ways.

The DAF was created to allow all universities and colleges to conduct a data curation and preservation audit of departmental data collections, awareness, policies, and practices. DAF, in principle, enables companies to collect information about their research data sets efficiently. This feature made this model an unavoidable inclusion in this study since it emphasized the need to understand the available research data sets, their location, how they are managed, and how they are stored.

The CCMF was created as a self-evaluation tool for institutions, funders, and evaluating researchers' capacity to conduct data-intensive research (Lyon *et al*. 2012: 126). Openness, expertise, training, technical infrastructure, and restrictions imposed by law and policy, as described by Cox and Pinfield (2014: 299), all play a critical role in the life-cycle of research data, from collection to assessment and retention, access and re-purposing.

The UCRDMF model was created to represent a paradigm shift in research data management system design by concentrating on user demands and the unique needs for research data needed to build effective systems. The UCRDMF depicts the creation process for a straightforward research prototype that considers a data management system using information gathering, prototype design, and system evaluation, making it a critical model for the study.

2.8 Appraisal of the Chapter

Chapter two provided a well-crafted and comprehensive exploration of the theoretical foundations that underpinned the study. The chapter plunged into an in-depth discussion of three distinct theoretical frameworks: the Data Audit Framework, the Community Capability Model Framework, and the User-Centered Research Data Management Framework. The thorough exploration was a fundamental building block for the subsequent chapters and set a solid academic footing for the research.

One of the chapter's notable strengths was the clarity and detail with which the theoretical frameworks were presented. The study provided a clear exposition of each framework, explaining its fundamental concepts and principles and how it aligned with the study's objectives. This enabled a grasp of the research's theoretical underpinnings and an understanding of how these frameworks contributed to the study's structure and direction.

The chapter also effectively elucidated the selection criteria for the chosen theoretical models. The study outlined the strategies used to identify essential variables and their complementary value to the research, demonstrating a thoughtful and deliberated approach to model selection. Additionally, including studies that have adopted these models in various institutions added credibility and relevance to the discussion, showcasing real-world applications and implementations.

CHAPTER THREE: LITERATURE REVIEW

3.1 Introduction

The foundation of a review of literature is an in-depth examination of research studies and theoretical arguments concerning a specific phenomenon or phenomena. A comprehensive literature review is a well-thought-out argument that proposes an explicit world state for communication researchers. The goals of a literature review should be to collect and synthesize material; justify why a specific study or review should be conducted; establish a theoretical stance; and acknowledge and account for both opposing and corroborating viewpoints (McEwan 2018: 2). There appears to be no uniform standard dictating how literature is organized, which is mainly a decision made by the researcher. The researcher can arrange the literature according to the study problem's broader themes, thematic areas deriving from research questions, and technique (Card 2012: 728; Torraco 2016: 404; Winchester & Salji 2016: 308).

This chapter's literature review is structured around the research objectives, themes, and underlying theory's crucial variables of the research problem.

The first section concentrated on analyzing the most pertinent studies on research repositories at HEIs using the systematic analysis approach.

The second section discussed ascertaining the management of the existing research for researchers. The section covered broad topics such as research data, creation, evaluation, preservation, disposal, and storage facilities.

The third section reviewed literature that sought to establish the usefulness of the faculty research data repositories. This section addressed significant issues such as research data sharing and reuse by researchers.

The fourth section examined the existing RDRs, design and development, metadata standards, and interoperability to inform the research objective that sought to create a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence.

The fifth section summarizes the primary existing RDM challenges obtained from the literature.

3.2 Systematic review analysis of the research repositories at HEIs

In research-based HEIs, measuring research performance has evolved into a systematic practice (Kanngieser, Neilson, & Rossiter 2014: 302). As a result, researchers and academics are more conscious of the need to increase each piece of research's value and effect. A systematic literature search was conducted on several databases, including Ebsco Host, Emerald, Elsevier ScienceDirect, Sage, Google Scholar, SABINET SA e-Publications, and abstract index databases such as Scopus and Web of Science to analyze the most relevant studies on research repositories at HEIs. The comprehensive review followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) standards. When performing the systematic review, the PRISMA standards were adhered to, which many academics have utilized (Moher et al. 2010: 5; Zibani *et al.* 2021: 239). PRISMA is an unambiguous set of items for writing reports in meta-analyses and systematic reviews grounded in research (Sam, Naicker, & Rajkoomar 2020: 609). Three key topics were devised to cover the entire scope of this systematic review and gain knowledge on the early adoption, growth, and advantages of faculty repositories (Zibani *et al.* 2021: 239). The initial query focused on the driving forces for creating academic research repositories. The second query covered the advantages of faculty research repositories and their services. The final query focused on how faculty research repositories were used.

3.2.1 Review synthesis

This section reviews the eleven pertinent full-text articles relevant to the analysis scope in the sequence of their appearance in Table 3.1. Additional summaries of these publications are provided in the section, organized according to their main ideas and contributions to research repositories.

a) **Establishing an open-access repository for Doctor of Nursing Practice projects**

To advance nursing practice and overcome the difficulty of dissemination by identifying the right audiences to guarantee extreme bearing, a digital repository for Doctor of Nursing

Practice (DNP) project work was created to document and exchange students' final DNP projects. A thorough analysis of the methods and organizational frameworks used by other institutions of a similar nature for creating repositories was carried out. DNP projects were effectively uploaded to the project repository. The audience map on the repository page indicated the project's impact on nursing practices worldwide. Their research findings were used in practice worldwide, accessed, referenced, and applied by others, motivating the content creators.

a) The pediatric imaging, neurocognition, and genetics (PING) data repository

This research discussed the establishment of a repository of uniformly measured interactive genetic traits, whole genetic material sequencing, and image analysis from children with typical development aged 3 to 20 years (Jernigan *et al.* 2016: 1151). The repository produced shareable data on the brain and cognitive development of 1493 children. The repository offered an intellectual data examination tool to make applying advanced statistical simulations to PING data easier, map effects by region of interest and vertex, and visualize the findings in three dimensions on the cortical surface. Additionally, it was discovered that the portal included suitable statistical demonstrating competencies, designed descriptors for every PING measurement, and an easy-to-use user interface to manage a variety of quality control and analysis workflows (Jernigan *et al.* 2016: 1151).

b) Building an archaeological data repository: a digital library and digital humanities collaboration at the University of South Florida

The study showed a real-world example of how the University of South Florida (USF) Libraries Digital Collections collaborated with the Institute of Advanced Study of Culture and the Environment (IASCE) to support the discoverability and accessibility of a new archaeological dataset in partnership with a campus-based Digital Humanities research facility (Mi, Bernardy & Schmidt 2021: 140). The Andean Archaeological Data Project, a prototype for an archive of archaeological data, was created along with a digital collection that was successfully housed in an established digital library platform, enabling public utilization of the material via the library's web pages. The project was successful in helping IACSE with its grant applications and ongoing research by providing a publicly available functionality by which archaeological data could be viewed and utilized by IASCE scientists and outside researchers (Mi, Bernardy, & Schmidt 2021: 141).

c) Analysis of research data repositories in India

This study examined and assessed the Indian research repositories listed on the Registry of Research Data Repositories (Re3data.org) regarding their types, subject areas covered, indexed content categories, software tools, standards, and implementation guidelines. The study used the content analysis method to analyze particular research data repositories in India. The research helped identify active, high-quality research data repositories in India and paved the way for their efficient development across various disciplines in India (Manu *et al.* 2018: 318).

d) Occupational hygiene-related research in South Africa: development of a research repository

This study focused on discussions held by a select group of individuals involved in Occupational Health (OH) research, government, and business in Gauteng who met to discuss creating a repository for OH research with a particular emphasis on occupational hygiene (Brouwer & Du Plessis 2016: 34). It was decided that the repository should contain students' research papers, technical reports, and scientific articles. A form asking for information about the occupational hygiene- and occupational health-related research carried out in the years (2013–2016) was distributed to identified researchers in universities and research institutions. The form asked for metadata about the research topic, a brief description, the research type (e.g., MSc, Ph.D., or other), the dissemination type, and full references. The repository was designed to give context for discussions about a South African occupational health and hygiene research plan by revealing the areas of current research being pursued and the main interests of research teams (Brouwer & Du Plessis 2016: 34).

e) A framework for managing E-records in higher education institutions: a case study of the Institute of Development Management, Eswatini

This study focused on establishing a concept model for efficient electronic records management at the Institute of Development Management (IDM) in Eswatini. To achieve this, the institute's readiness for electronic records was examined (Tsabedze 2019: 4). The analytical structure of the research was grounded on the e-records readiness assessment tool. The study's findings showed that despite IDM relying heavily on computers to manage information and communications, the organization's registration system was not properly managed to facilitate its commercial operations in support of its business operations (Tsabedze 2019: 7). Institutional policies and guidelines that would have guided the management of electronic records, as well as a lack of knowledge and expertise in managing electronic records, were amongst the factors contributing to this neglect (Tsabedze 2019: 11).

f) Faculty use of a learning object repository in higher education

This study aimed to understand how instructors employ learning object repositories (LORs). Faculty users of a Wisconsin Online Resource Center LOR served as the case study's subjects (Xu 2016: 471). Ninety-two participants in the study completed a survey, and the data were examined with descriptive statistics and a Fisher exact test. The study discovered crucial avenues by which faculty members were informed of LORs. The study advanced LOR design and service from the angles of advertising, content creation, integration with systems for managing instruction, and assistance with technology (Xu 2016: 471).

g) Supporting nursing faculty with a digital repository of simulation resources

A nursing school with multiple campuses encouraged faculty members to use simulation by building a digital library of modeling resources adapted to faculty needs (O'Neill *et al.* 2020: 175). In this descriptive empirical investigation of faculty first perceptions and repository utilization, five patterns were identified: an appreciated resource; one that is simple to use and access, one that promotes consistency across campuses, one that helps new simulation educators, and one that increases the use of simulation pedagogy. A digital archive of simulation resources is a cutting-edge tool for inspiring modelling use amongst nursing professionals. The exploration team first evaluated fourteen modelling websites that had been shortlisted and were categorized as resources, materials, and educational products to develop the digital archive. The intended community welcomed the resource, and usage also reportedly increased since the repository's creation (O'Neill *et al*. 2020: 175).

h) Open access repositories on open educational resources: feasibility of adopting the Japanese model for academic libraries

According to Leng, Ali, and Hoo (2016: 37), the Wawasan Open University Library launched a research project to establish open-access repositories for cost-free educational materials. The study showed how two OER web portal archives were built using the Japanese open-source program WEKO. The design process followed a pull-to-push methodology in which the Open Archives Initiatives Protocol for Metadata Harvesting method was used to collect the record of academic open-access materials kept in the institution's and network sections of the community's online libraries (Leng *et al.* 2016: 37). The collected metadata was then uploaded to another open knowledge platform for other users to find. The development of university open-access repositories produced results that effectively made the content from this open knowledge platform freely available for reuse while strengthening the librarian's position as manager of institutional assets.

i) The MTA SZTAKI smart factory: a platform for research and project-oriented skills development in higher education

The study described a "Smart Factory" platform offering students tasks and materials for developing project-oriented skills. The platform connected these opportunities to specialized higher education by holding individualized student projects and classes with a predetermined progression (Kem'eny *et al.* 2016: 58). It was demonstrated that fewer students could still gain important knowledge and practical experience, both through individual projects and structured coursework, even though the facility's dimensions and capabilities were nowhere near those of a full-fledged learning factory (Kem'eny *et al*. 2016: 58).

j) Indian Research Information Network System (IRINS): An Analysis of Faculty Profiles of The Gandhigram Rural Institute - Deeded to the University

The Indian Research Information Network System (IRINS), a Reference Information Management (RIM) created by the Central University of Punjab, was used in the study to review academic articles and different departments publications with citations of the Gandhigram Rural Institute (Tamizhchelvan & Anbalagan 2020: 9). The results showed trends in direct citations increasing, faculty members' average citation counts in publications varying between CrossRef, h-index, and Google Scholar Citations, and four faculties had faculty members ranking in the top ten (Tamizhchelvan & Anbalagan 2020: 9).

The above articles provide rich insights into establishing and utilizing research repositories in various academic and professional contexts. Each article presents a unique perspective on how digital repositories contribute to the dissemination, accessibility, and management of research data and knowledge. A synthesis of these articles, categorized based on their core themes and contributions to the realm of research repositories, is provided below:

Establishing Open-Access Repositories:

The first article emphasizes the creation of a digital repository for Doctor of Nursing Practice (DNP) projects to enhance nursing practice and ensure widespread dissemination to pertinent audiences. The repository successfully documents and shares DNP projects, facilitating global access and utilization, thus advocating for the significance of open-access repositories in specialized fields.

Specialized Data Repositories:

The second and seventh articles delve into establishing specialized data repositories for pediatric imaging and nursing simulation resources, respectively. These repositories demonstrate the importance of tailoring repositories to specific domains, ensuring efficient access to crucial data and resources.

Archaeological and Occupational Data Repositories:

The third and fourth articles discuss the establishment of repositories in archaeology and occupational hygiene. These repositories aim to organize and disseminate research data, showcasing the significance of repositories in domains where systematic data organization is pivotal for future research and policy formulation.

Country-Specific Research Data Repositories:

The fifth article focuses on analyzing research data repositories in India, highlighting the importance of evaluating country-specific repositories to enhance the quality and efficiency of research data management and access.

Educational and Learning Object Repositories:

The sixth and eighth articles shed light on using repositories in the educational context, emphasizing learning object repositories and simulation resources. These articles underscore the potential of repositories in supporting educational endeavors, aiding content creation, and enhancing pedagogical practices.

Institutional Repository Development:

The ninth article discusses the development of an innovative factory platform that supports project-oriented skills development in higher education. This article emphasizes the role of repositories in educational institutions to facilitate practical skill development through structured coursework and projects.

Research Information Network and Profiling:

The final article uses an academic profiling system, the Indian Research Information Network System (IRINS), to analyze faculty profiles and academic publications. This article underscores the role of repositories in academic information management and profiling.

These articles collectively demonstrated repositories' diverse applications and importance in organizing, sharing, and utilizing research data and knowledge across various domains. From specialized data repositories to educational repositories and country-specific data repositories, these studies showcased the versatility and potential impact of well-structured repositories in advancing research, education, and societal development.

Table 3.1. List of articles included in the systematic review of faculty research repositories in higher education institutions

ID	Author	Title	Year Country	Type
A1	Heselden, M. et al.	"Establishing an open-access repository for doctor of nursing practice projects."	2019 USA	Conceptual
A2	Jernigan, T.L. et al.	"The pediatric imaging, neurocognition, and genetics (PING) data repository."	2016 USA	Experimental

A3	Mi, X., Bernardy, R., Schmidt, L.	Building an archaeological data repository: a digital library and digital humanities collaboration at the University of South Florida."	2021 USA	Conceptual
A4	Manu et al.	"Analysis of research data repositories in India"	2018 INDIA	Conceptual
A5	Brouwer, D. du Plessis, J.L.	"Occupational hygiene-related research in South Africa: development of a research repository."	2016 SOUTH AFRICA	Conceptual
A6	Tsabedze, V.	"A framework for the management of E-records in higher education institutions: a case study of the Institute of Development Management, Eswatini."	2019 SWAZILAND	Conceptual
A7	Xu, H.	"Faculty use of a learning object repository in higher education"	2016 USA	Conceptual
A8	O'Neill, B. et al.	"Supporting nursing faculty with a digital repository of simulation resources."	2020 AUSTRALIA	Experimental
A9	Leng, C. B., Ali, K. M. and Hoo, C. E.	"Open access repositories on open educational resources: feasibility of adopting the Japanese model for academic libraries."	2016 MALAYSIA	Experimental
A10	Kem'eny, Z. et al.	"The MTA SZTAKI smart factory: platform for	2016 HUNGARY	Experimental

		research and project-oriented skill development in higher education.”		
A11	Tamizhchelvan, M. and Anbalagan, M	“Indian Research Information Network System (IRINS): an analysis of faculty profiles of the Gandhigram Rural Institute – Deeded to be University.”	2020 INDIA	Conceptual

Source: Zibani *et al*. (2021)

3.3 Research data

Fundamental research has long-valued research data, which is not a brand-new concept. University research data management guidelines list various widely accepted definitions of research data. According to the Queensland University of Technology Management of Research Data Policy (2016), research data are pieces of information that support a claim, theory, test, hypothesis, or other research output. Facts, observations, pictures, photographs, recordings, measurements, or experiences can all be included in this material. Any format or medium, including raw, cleaned, and processed, as well as numerical, descriptive, visual or tactile, can be used to store research data (CARL Data Management Sub-Committee 2009: 4). According to the CARL Findings Management Sub-Committee, research data records are unanimously regarded in the research community as crucial for validating scientific data (2009: 4). Research data has gained popularity due to its vital role in contemporary scientific discoveries and Scientists are now sharing it as a fundamental component of scientific growth. According to the National Science Board (2005: 18) and Thomas (2011: 37), following the technique of data capture can be divided into four categories: Observational (data collected by a researcher assertively actively participating in inducing and assessing change or in having an impact when an attribute is transformed); Simulational (data are obtained by simulating the performance of a process or framework in real life); Derived (data are obtained by extrapolating from the results of an experiment); and Experimental (data involves using current data statements, often from various sources of data, to develop new data through some conversion, like a mathematical model or grouping). Scholars such as Krier and Strasser (2014: 29) categorized data based on forms such as qualitative (non-numerical with variables that can be nominal or ordinal) and quantitative data (numerical with variables that can be continuous or discrete) (Ohaji 2016: 25). Data can also be divided into categories based on the primary types of the subject areas (Bangani & Moyo 2019: 12).

3.3.1 Management of research data

The loss of research data is a common occurrence in research. Scholars such as Velden and Lagoze (2009: 675) and Briney, Coates, and Goben (2020: 1) contend that data have a lengthy history of being unavailable when needed by researchers. For example, Pearlman (2009:1) discovered that NASA's historically significant scientific data was probably wiped out in the 1970s magnetic tape shortage. Vines *et al.* (2014: 95) found that 17% of biological data perish annually after a study is published. (Briney *et al*. 2020: 1) asserts that any researcher can benefit from well-managed data because it requires less searching, less effort to understand, and less processing to get ready for sharing, reuse, and collaboration. A hard-drive crash or losing a crucial companion does not have to stop a project or necessitate data collection.

On the other hand, the fundamental motive for data management for research is that it helps the researcher focus on scientific challenges rather than data administration (Katabalwa, Bates, & Abbott 2021: 6). This section covered the essential research data management strategies researchers use across all fields. This section addressed some commonly used terms and methods, highlighting the instances where they intersect with research data management. These concepts and procedures included research data, the life-cycle of research data, and models. Although some terminologies were discussed in Chapter One, this section gave a more in-depth examination.

3.3.2 Research data lif-ecycle

Many models have been developed to show the various stages of research data existence. These models are demonstrated after a specific attribute, such as a particular subject of study or to serve a specific goal; from a generic reference; or research data management in general (Corti *et al*. 2014: 75). These models depict a more extensive system's life-cycle, such as the research process, as a sequence of related stages or phases in which data is generated or processed (Ball 2012: 11). According to Carlson *et al.* (2011: 645), research data life-cycle models help graphically identify and depict complicated processes, pinpointing crucial portions or various stages of existing research and the accountable individuals or organizations. There are several research data life-cycle models, each with a unique emphasis or technique. Carlson *et al*. (2011:

645) categorize research data lifecycle models according to their shape (linear, circular, non-linear, or other models) or setting (individual-based, organization-based, and community-based). Although the current project's primary purpose is to design a research data repository platform conducive to a University of Technology, encompassing research data storage, data access, and data reuse, it also touches on other themes such as research data attributes and management methods. Several research data life-cycle models have been explored, praised, and employed by academics in their research pursuits because they are critical in illustrating various phases of data creation. This study reviewed the research data life-cycle models to guide how research data is created and managed.

3.3.2.1 Data Curation Centre Curation Life-cycle Model

The Data Curation Centre (DCC) Curation Lifecycle Model is a community-based lifespan model that defines the many steps of data curation thoroughly addressed but not integrated into the value chain of a research study (Bugaje 2019: 27). Data curation and preservation are highlighted in the DCC, which can be used to plan data management tasks. Most activities and stages described in the model are used in various research data repositories. Research data repositories bear a large portion of the load in the outermost ring where data storage, access, usage, and reuse are highlighted (Bugaje 2019: 27). Research data repositories are crucial for community involvement and watch because they provide a hub for data creators, users, and professionals (see inner ring). This model provides a comprehensive graphic summary of the life-cycle phases necessary for practical preservation (Higgins 2008: 136). Various scholars (Ng'eno 2018: 51; Shakeri 2013: 15; Chawinga 2019: 65) have used this model.

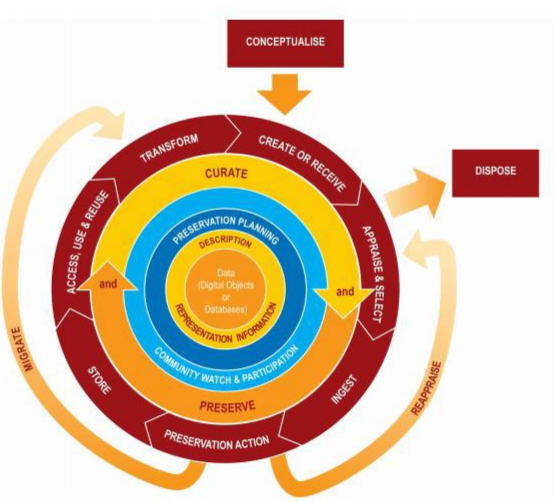

Figure 3.1. DCC Curation Life-cycle Model

Source: Higgins, (2008)

3.3.2.2 Open Archival Information System (OAIS)

The International Organization for Standardization (ISO) created the OAIS theoretical foundation for guiding the development of processes for the long-term preservation and conservation of information (Consultative Committee for Space Data Systems 2012: 2). This is one of the most popular models for dealing with the issues surrounding the maintenance of digital artifacts over time (Higgins 2008: 136; Schumacher & Vande Creek 2015: 99; Jeng, He, & Chi 2017: 626; Schöpfel *et al.* 2018: 248.). The study by Jeng *et al.* (2017: 626) examined the methods and difficulties of data curation within the OAIS paradigm, an area where independent research was noticeably lacking. The researchers aimed to close this gap by asking two important research questions. The Interuniversity Consortium for Political and Social Research (ICPSR), the world's largest social science data repository, was the focus of the first study question, which sought to understand the practices currently used within a data repository. The second study question sought to understand the technological challenges data repositories encounter and the IT solutions that were crucial to enhancing their data repository services. The study mapped collected information to the OAIS framework using a case study methodology focusing on ICPSR to assess alignment.

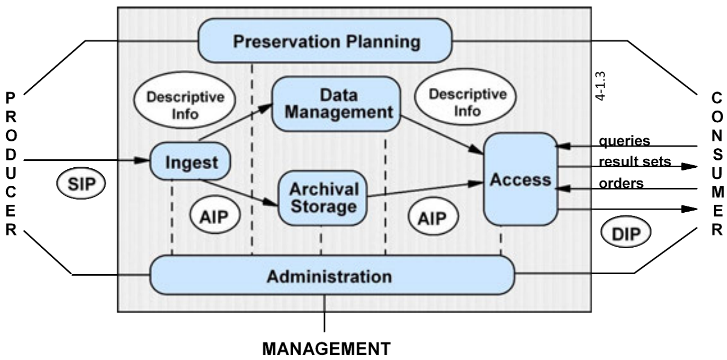

Figure 3.2. The OAIS Model

Source: Consultative Committee for Space Data Systems, (2012)

3.3.2.3 United States Geological Survey (USGS) Science Data Lifecycle Model

According to Faundeen (2010: 21), the model explains the steps involved in a proposed research project and the activities carried out, such as conception, gathering, processing, evaluation, archiving, dissemination, and data sharing for re-purposing. Some actions should be done regularly throughout the life cycle, such as documenting the workflow process, providing metadata, and backing up data to prevent physical loss. Various researchers (Chawinga 2019: 65; Pouchard 2015: 342; Faundeen 2010: 22) praised and embraced this concept.

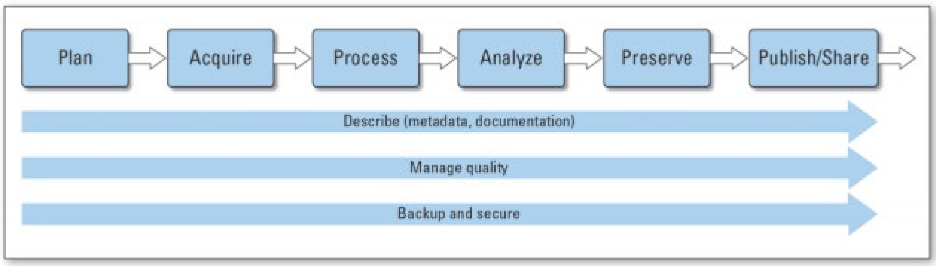

Figure 3.3. USGS Science Data Life-cycle Model

Source: Faundeen *et al.* (2010)

3.3.2.4 The Community Capability Model Framework

Lyon *et al*. (2012: 130) developed this model as an evaluation tool for discipline researchers, including eight competence elements representing human, technical, and environmental challenges. The model has been frequently used in academic research (Lyon *et al*. 2012: 130; Qin, Crowston & Lyon, 2016: 25; Chawinga 2019: 65). To increase the effectiveness and efficiency of RDM, the CCMF serves as a viewpoint tool for organizations, funders and scholars. It outlines the roles, responsibilities, and requirements for each capacity. Since data curation concerns digital research data, this is particularly significant in faculty-generated research data.

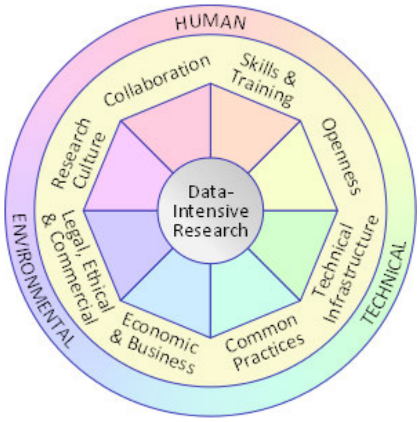

Figure 3.4. The Community Capability Model Framework

Source: Lyon *et al*. (2012)

3.3.2.5 UKDA Research Data Life-cycle

The UKDA Research Data Life-cycle model examines the span of an aggressively utilized dataset that helps compare how data management operations connect to different stages of a research endeavor. This model is similar enough to other life-cycle models to be a fair depiction and has been used in various studies (Nhendodzashe 2017: 75; Yunus 2017: 19).

Figure 3.5.Data life-cycle model

Source: UK Data Archive, (2017).

3.3.3 Research data creation

Research data creation is a crucial step in ascertaining the management of data. Steps comprise developing a data management plan, obtaining permission for sharing, identifying already-existing data, gathering data, creating a method for data collection, and producing metadata (Van den Eynden 2013: 21; Van Wyk & Van der Walt 2014: 10; European Commission 2016: 7). Research data creation is defined by the European Commission 2016: 7) as the process of gathering data that will be processed and used later to meet specific aims. It involves digital or paper information that is research-related data collected, produced, or acquired. It is successively used as a basis for making analyses or making judgments by the researcher to advance, endorse, or modify hypotheses, processes, and research results (European Commission 2016: 7). According to Edinburgh University Library (2011: 6), when creating a research project, a researcher must consider which file formats to use for data storage. In other cases, file formats are established by the software being used or by the standards of the discipline. The Edinburgh University Data Library Research Data Management Handbook (2011: 6) gives a list of research data formats, which includes, but is not limited to, the following:

i. Text files (MS Word documents,.txt files, PDF, RTF, and XML) (Extensible Markup Language);
ii. Numeric - SPSS, Stata, and Excel;
iii. Multimedia - jpg/jpeg, gif, tiff, png, MPEG, mp4, QuickTime;
iv. 3D and statistical models;
v. Programming languages - Java, C, Python; Subject-specific formats - Astronomy's Flexible Image Transport System (FITS), crystallography's Crystallographic Information File (CIF); and
vi. Specific instrument formats - Olympus Confocal Microscope Data Format.

Scott (2014: 121) suggests that the following research records may be valuable to manage both throughout and after the project's life cycle:

i. Electronic mail and print correspondence;

ii. project files;

iii. funding requests;

iv. ethical requests;

v. technical documents;

vi. research report; and checklists; and

vii. Forms of consent that have been signed.

3.3.4 Research data description

According to the Southeastern Universities Research Association (SURA) (2012: 3), extensive data documentation is necessary for future data comprehension. "Data cannot be easily located, understood, or efficiently used without a thorough explanation of the context of the data file; the setting in which the data were collected; the measurements that were taken; and the quality of the data" (SURA 2012: 3). In support of this statement, Patel (2016: 240) posits that accurate data retrieval necessitates data description and identification, which can only be achieved using a specialized metadata format for data representation. The usage of an appropriate metadata standard is required for data description. If one provides detailed information, others can find, interpret, and use one's data (Patel 2016: 240). The discipline or community most pertinent to the research should create metadata in a widely accepted format (Kabanda *et al.* 2023: 9). Data descriptions should comprise a data dictionary, a defined data model, a summary of the research endeavor, descriptive file names, standard terminology to aid finding, spatial location, space and time formats, and a summary of the contents of data files (SURA 2012: 3).

3.3.5 Identification of research data

To ascertain data management, universities, and research institutions should have suitable criteria for data identification, classification, and preservation to fully utilize the potential of research data. An institutional strategy for identifying, classifying, and locating research data can raise awareness of collection strengths and difficulties that need to be addressed (Jones *et al.* 2009: 5). An institution with suitable processes for identifying, classifying, and locating research data sets is well-positioned to maximize the value of its research data collections while allowing ongoing use and through literature review. Institutions have utilized various methods to identify research data in their possession.

3.3.5.1 Data Management Plans

Although RDM has always been an essential skill for researchers and librarians, introducing open data guidelines and funding organizations' directives, have resulted in a considerable body of new knowledge practices around research data (Antell *et al*. 2014: 566). Data Management Plans (DMPs) have become a standard component of the research process and grant applications in response to appeals for enhanced digital data administration generated by sponsored research initiatives worldwide (Mischo, Schlembach & O'Donnell 2014: 37; Whitmire, Boock, & Sutton 2015: 385). The European Commission's Horizon 2020 project defines DMPs as critical elements of successful data management. A DMP outlines the data management workflow that a research project will gather, handle, and/or generate (European Commission 2016: 3). As an undertaking to make research data findable, accessible, interoperable, and reusable (FAIR), A "DMP should contain information on:

- How research data will be handled throughout and after the project's conclusion;
- What data will be gathered, processed, and/or created
- What criteria and procedures will be applied?
- If data will be shared or made accessible to everybody; and
- How data will be organized and maintained after the project is finished." (European Commission 2016: 3).

In essence, DMPs are typically brief, advanced plans for description that outline the data a research project will produce; how it will be stored (securely, if necessary); who will have access to it; and what records and metadata will be generated with it; as well as archival measures if the data are to be kept for a long time (Mischo *et al.* 2014: 37; Whitmire *et al*. 2015: 385).

3.3.5.2 RDM policy frameworks

Policy frameworks serve as a basis for the proper management of research data. According to the Office of Scholarship at the University of Cambridge RDM policy framework (2021), an

RDM policy framework aims to advise research staff and students on creating, managing, and sharing research data to optimize research potential. Similarly, Patel (2016: 241) asserted that an institutional RDM legal and policy framework must define the goals, parameters, areas of use, and instructions for data sources. Important topics, including licensing, metadata, data classification, copyright agreements and conditions, terms and conditions for using data, protecting the confidentiality of sensitive data, and data protection, must be covered from security breaches and concerns about intellectual property in the context of RDM. Therefore, RDM frameworks address critical challenges in managing research data, including data protection, confidentiality, ownership, and licensing, by clearly presenting the workflow of the data life-cycle in its different phases, from generation to storage, organization and sharing (Heffetz & Ligett 2014: 80). The presence of RDM frameworks within institutions leads to establishing and implementing proper and effective research data processes ranging from research data capture to evaluative, secure, open, re-purposed and shared (Cox & Pinfield 2014: 306).

3.3.5.3 Data audit assessment

Ascertaining data management entails knowing what is held and by whom to identify duplication of effort and prioritize resources. Jones *et al.* (2009: 5) posit that although higher education generates vast amounts of data, few institutions have formal policies for the long-term curation of these research data. Therefore, data auditing can provide several advantages to an institution, such as cost-cutting, risk management, and facilitating access and reuse. These advantages depend on understanding research data holdings (Patterton 2016: 242). Knowing how data is selected and whether or not controls are in place will reveal potential risk areas.

In the same way, knowing research data agreements is critical for facilitating access and encouraging reuse. An understanding of research data holdings is essential for good data management. The previous chapters have elaborated on the considerable uptake of audit assessments by various global institutions such as Bath, Edinburgh, Glasgow, Ghana, NUST (Zimbabwe), research institutes, and CSIR(South Africa) (Jones *et al.* 2009: 8).

3.3.5.4 Research data storage

As Kostkova *et al.* (2016: 2) state, research data is an essential science commodity. Therefore, proper management is critical. It is also important to note that scientific data has significantly changed with the quick pace of technology. This change has led to the emergence of various data file types, dataset sizes, data complexity, use cases, and data-sharing techniques. While all forms of data are essential, data curation procedures must evolve to keep up with the changing nature of all types of data and new data collecting and storage methods (Kostkova *et al*. 2016: 2). Research data storage is a critical aspect of data management as it allows researchers to save digital material for retrieval at a later stage. Researchers' responsibility for securely maintaining data has become more important and challenging, especially as their data collection expands. Researchers can no longer write data on an external disk and keep it in their filing system or save it to their computer systems for sharing. Kostkova *et al*. (2016: 2) present ten basic guidelines for digital data preservation based on talks amongst teachers for the Software and Data Carpentry initiatives, namely:

- Anticipate how the data will be used- a clear plan of what to expect before data collection begins is essential.
- Know one's use case - clearly understanding one's use case makes data storage much more accessible.
- Keep raw data raw - access to the raw data can facilitate future review and diagnostic replicability.
- Keep data in open formats - to improve access and significance over time, and data should be stored in formats with open conditions.
- Data should be structured for analysis – to allow easy comprehension and analysis.
- Data should be distinctively recognizable - with unique identifiers, such as Digital Object Identifiers (DOI)
- Link relevant metadata - metadata must be well-defined and firmly linked with data.
- Adopt the proper privacy protocols – ensure one's plans to secure data's confidentiality.
- Have a systematic backup scheme - at every stage of the research process.
- How much data one has determines where and how it is stored - the storage technique used is determined by the scope and nature of the data and the escalating expense of access and storage.

3.3.5.5 Appraisal of research data

Research data cannot be kept indefinitely. Therefore, appraisal processes must be in place in research-intensive institutions. Whyte and Wilson (2010: 1) suggest that while all digital data cannot be stored forever, there is no broad acknowledgment of the necessity to curate data outside the archive and library communities. Instead, some people believe storage is inexpensive and want to keep everything. Whyte and Wilson (2010: 1) further state that while it may be technically feasible in theory, there are four significant challenges to this viewpoint:

- ***The growth in the amount of digital content***. "...if the expansion of content (per byte or object) keeps pace with the lowering cost [of storage], the true cost of preserving everything may be the same as it is now, or even greater."

- ***Costs rise as a result of backup and mirroring***. Without suitable mirroring and backup technologies, no digital preservation strategy can last, increasing the cost of storage by at least a factor of two.

- ***The task of discovery becomes more difficult***. Keeping everything results in searches with a high clutter, necessitating more work to determine whether data is the intended target of an investigation.

- ***Maintaining and sustaining a collection is costly***. It is necessary to factor in the cost of developing and maintaining maintenance metadata and the expense of performing preservation actions on data that must be kept.

Research data appraisal and selection are essential steps in data management processes (Van den Eynden 2013: 10; Van Wyk and Van der Walt 2014: 21). It entails a review of research data to determine which to keep indefinitely, which to keep for the time being, and which to reject (Higgins 2008: 134). Appraisal has been regarded as one of the most significant responsibilities in RDM due to the vast amounts of research data that research institutions produce. Furthermore, when analyzing research data and selecting those that require long-term preservation, it is necessary to follow established guidelines, rules, and regulatory criteria. Appraisal and selection policies must ensure that decisions are made consistently, transparently, and responsibly. According to Whyte and Wilson (2010:3), evaluation and selection procedures must comply with legal requirements such as privacy and intellectual property rights, Public Records Acts, national data policies, and codes of conduct issued by the research institution or funders.

3.3.5.6 Selection and Appraisal Policy

It is pertinent that research institutions have established policies for selecting and evaluating research data. Similarly, Whyte and Wilson (2010: 3) reveal that the selection and appraisal policy includes standards for assessing the value of a research dataset and what should be done with it in response, such as when it can be destroyed or how long it should be kept. Depending on whether the mandate covers preservation, different standards apply. In any event, the policy serves as a foundation for subsequent data analysis. This is driven by discipline-specific traits and based on general standards like the seven listed below, which were gathered from various sources (such as NASA Socioeconomic Data and Applications Center 2010: 28; Faundeen 2010: 17):

- Relevance to Mission: The resource material is relevant to the center's mission and any priority specified in the current plan of the funding body or research organization, including any legal requirements to keep the data after it is used.
- Importance in Science or History: Is knowledge valuable in science, society, or culture? To evaluate this, it is necessary to infer expected future use from current research evidence and educational value.
- Exclusivity: To what extent is the source the exclusive or best source of information gleaned from it, and is it in danger of being lost if it is rejected, or can it be retained elsewhere?
- Prospective for Redistribution: The data files' reliability, validity, and applicability can be assessed; they are received in formats that adhere to strict technological specifications; and any issues with intellectual property or human subjects can be rectified.
- Non-Replicability: Reproducing the data or resource would be impossible, unfeasible, or expensive.
- Economic Argument: Managing and protecting the resource's costs can be determined, and they are reasonable compared to the evidence of potential future benefits; money has been procured if necessary.

- Complete documentation: all data required for future discovery, access, and reuse is accurate and complete, including origin metadata.

As a result, researchers, librarians, archivists, and IT specialists in research institutions must undertake appraisals to determine whether research data should be acquired for long-term or short-term preservation. Consequently, this process enforces a more significant requirement to determine retention periods per the institution's policy frameworks. According to Whyte and Wilson (2010: 3), research data librarians and archivists should define assessment/selection criteria and appraisal policy with input from stakeholders, in particular, researchers and local data managers, as it is crucial to understand how data will be assessed and how to maximize their long-term effect.

3.3.5.7 Research data preservation

Research data preservation as a data management process comprises transferring data to the most appropriate format and media; backing up and storing data; creating metadata and documentation; and archiving data (Van den Eynden 2013: 10; Van Wyk and Van der Walt 2014: 21). It involves ensuring permanent access to the original research data generated by the finished research project, as well as data accessibility to others for data verification, exchange or collaboration within the scientific community. Hence, crucial components of data preservation operations in universities and research institutes include long-term preservation and protection of sensitive data. Universities and research institutions should have guidelines stipulating where to store research data, clearly define the retention periods, and handle the discrepancies between quick access and long-term safeguarding, depending on the type of research data (Chiware and Mathe 2015: 5; Matlatse 2016: 78). Planning should be done to cover administrative processes needed before beginning preservation efforts as well as the technical specifications for preservation. These preservation measures should guarantee that the research data from the university or research institution are legitimate and trustworthy while keeping their integrity.

3.3.5.8 Disposal of research data

Data management requires proper procedures for it to be disposed of efficiently. The US Department of Education (2014: 1) established the Privacy Technical Assistance Center (PTAC) and noted that research data disposal spans various media, including electronic and paper records. The risk created by the sensitivity of the data being destroyed, as well as the possible impact of unauthorized disclosure, should guide the manner of disposal chosen. For instance, the negative impact of disclosing a file containing directory information, such as honor roll students' names, may be less severe than disclosing a file including students' Social Security numbers, names, and dates of birth. As a result, the data disposal strategy in these two cases may differ. Research organizations progressively collect and maintain vast research data in their academic pursuits. As noted by the US Department of Education PTAC (2014: 2), some research data may need to be kept permanently, while others may need to be kept for a certain amount of time to meet legal or policy retention requirements. Once those periods have passed, such data will need to be erased.

The UK Data Service (n.d) gives essential information when discarding research data. For example, simply removing files and reformatting a hard disk will not securely remove information, meaning that data previously stored on the hard drive can still be recovered. A method for safely wiping data files is an integral part of securely maintaining data, and it can be helpful at different points of the data cycle. Deleting files from hard drives, which are magnetic storage devices, does not permanently delete the file. Rather, it eliminates a reference to the file. It takes little effort to recover files that have been accidentally destroyed in this manner, which explains why data can be restored from some damaged hard drives. Files must be organized. To ensure that files are effectively unreadable, they must be rewritten multiple times. The UK Data Service (n.d) further stipulates that memory sticks and other flash-based storage devices are built differently than hard disks. The only reliable method of erasing files is physical destruction because methods for securely wiping files on hard drives do not work for solid-state disks. Paper and optical media should be destroyed with shredders certified to an adequate security level.

In South Africa, the Protection of Personal Information Act, No. 4 of 2013 (POPI) was instituted to honor the right to privacy as enshrined in Section 14 of the Constitution of the Republic of South Africa (1996). The POPI Act distinctly outlines the protocols for retaining and discarding personal information even after the intended collection purpose timelines have elapsed.

3.4 The usefulness of faculty research data repositories

Scientific research has become increasingly data-driven and collaborative, posing more demands on faculty research data management techniques, and attitudes are scrutinized more frequently and intently. The need for evidence of more rigorous study and planning for research output comes from funding organizations and publishers, which has required an examination of existing faculty research repositories to determine the usefulness and benefits of their existence (Zibani *et al*. 2021: 239).

This is due to changing formats, the volume of collected data, and the emphasis on open access. Establishing faculty research data repositories is gradually gaining support because such initiatives have several advantages. After all, it frequently contains permanent faculty-produced research content. Although this content can be found elsewhere, for example, in the funder or publisher repositories, the faculty research repositories benefit from having data curated in a single location, widening appeal to a vast audience of study and instruction within the faculty (Wilkinson *et al*. 2016: 3). This section will provide an insight into the overall usefulness of faculty research data repositories.

3.4.1 Research data sharing and reuse

A wide range of research stakeholders, including research funders, journal publishers, and certain researchers, have demonstrated unmatched dedication to and interest in sharing research data. These parties have collaborated to make the world a data-sharing entity (Fecher *et al*. 2015: 18; Guedon 2015: 85; Matlatse 2016: 78; Chawinga 2019: 103). Research data-sharing is a complex and multifaceted concept that operates in the context of a spectrum of variables, some of which contribute to a thriving data-sharing ecosystem while others work against it. All research data management operations revolve around data-sharing (Wallis, Rolando & Borgman 2013: 12).

Sharing research data is crucial to achieving open science's core objectives, which include making science visible, reproducible, and accessible (Rousi 2023: 597). Research data sharing and FAIR data (findability, accessibility, interoperability, and reuse) are actively supported by academic libraries (Wilkinson *et al*. 2016; Cox *et al.* 2019). The European Commission's efforts to improve the FAIRness of all research data produced through projects with EU funding serve as one example (European Commission [EC], 2019). However, doing so can be difficult because different scientific domains face distinct obstacles as they advance toward sharing research data (Rousi 2023: 597). Researchers from subfields with different data-sharing cultures may work in research organizations. Without detailed information about the methods used by researchers at a given organization to acquire and distribute research data, it's possible that efforts to provide research data services won't be in line with their requirements. Multidisciplinary survey questions have frequently been used to examine the data-sharing practices of researchers. Many programs have been formed to encourage open data sharing and reuse, including Elixir, OpenAIRE, and the European Open Science Cloud (Katabalwa *et al.* 2021: 5). Additionally, campaigns like the Guardian's "Free Our Data" project forced the UK government to make 8000 open datasets available. The African Open Science Platform (Katabalwa *et al.* 2021: 5) promotes openness in African science.

Pro-data-sharing groups have sprouted across Europe, Africa, Asia, and other continents. Major national, regional, and worldwide organizations fund research initiatives and lead the data-sharing push. For example, the National Science Foundation of the United States of America (USA) has long had a policy requiring that researchers and research institutes receiving government funding deposit their findings in open-access repositories (Cohn 2012: 1004).

Moreover, the European Union (EU), a significant funder of research worldwide, requested that all data created by funded research programs be made freely available to the public beginning in 2014. (European Commission 2012: 3; Fecher *et al*. 2015: 18). In Africa, Chiware and Mathe (2015: 5), Koopman and de Jager (2016: 3) and Matlatse (2016: 78) reported that the NRF, a major funder of research in South Africa and parts of Southern Africa, required that all the research activities it funded have their resulting data made publicly available for reuse without any restrictions.

To promote data-sharing, reputable journal publishers have partnered with research funders to establish, implement, and promote rules requiring researchers to provide their research data. Journals like Atmospheric Chemistry and Physics, Nature, and PLOS One have policies requiring authors to submit their manuscripts with their research data (Fecher *et al*. 2015: 18). Such enthusiasm from significant research stakeholders is enough to suggest that novel ideas for the management and dissemination of research data are not just alluring but also realities that research stakeholders must accept.

Recently, Transformative Agreements (TAs) in the context of Open Access Publishing (OAP) have been the subject of an increase in conversations and announcements since the release of Plan S in 2018 (Zibani, Tomic, and Greene 2022: 2). Plan S is an initiative for open access publication that was started in September 2018 and is supported by cOAlition S, a global alliance of research financing and performing organizations. From 2021 onward, Plan S requires that scientific publications and underlying data from research supported by public grants be published in open-access journals or platforms that comply with its requirements. Numerous OAP agreements have been signed between institutions and publishers (libraries, national and regional consortia) due to the introduction of TAs through Plan S. The Efficiency and Standards for Article Charges (ESAC) Community claims that TAs between institutions and publishers reroute subscription costs to encourage open publication for authors connected to the negotiating institutions. The goal is to change the traditional subscription-based business model of scholarly journal publishing to one where publishers are paid over time (Zibani *et al.* 2022: 2). Numerous OAP agreements have been announced since then, with the Elsevier-UK Institutions Agreement being the most recent. The traditional subscription model's suitability for usage by research institutions has come under scrutiny in light of the rise of open access in scholarly journals in recent years (Zibani *et al.* 2022: 2).

According to Martone, Garcia-Castro, and Van den Boss (2018: 112), sharing data entails making actual data and any supplementary materials available to analyze data gathered for a research study. Research data sharing is not just about making research data available to the public, as several more research findings are highly valuable outside of the context of the original study. Sharing research data promotes new scientific inquiry, reduces data gathering repetition, and offers rich practical resources for instruction and training. Sharing data also promotes transdisciplinary research and learning and its discovery and use in fields other than the one in which it was created (Katabalwa *et al.* 2021: 5).

According to Jeng and He (2022: 1281), historically, professional groups in data curation and data management have relied on profile technologies to compile succinct and organized data about researchers and their research data. According to Jeng and He (2022: 1281), this profiling approach has been useful in examining data-sharing activities since it helps different stakeholders (such as institutions, discipline communities, and data repositories/centers) better understand researchers' actual and willing data-sharing behaviors. Research data sharing among researchers is hampered by several factors, including concerns about data misuse, a lack of technological infrastructure, a lack of technical expertise, inadequate funding, a lack of institutional and governmental support for data management, time constraints for submitting data to repositories, a perceived lack of personal benefits, and confidentiality (Bates, 2017; Katabalwa *et al.* 2021).

The POPI Act stands as a significant impediment to data sharing, particularly when there's potential for such sharing to infringe upon individuals' privacy (South Africa, 2013). The act delineates the circumstances in which data can or cannot be shared. It mandates that individuals whose personal information is being collected for research must be informed. Moreover, research participants have the right to object to sharing their data as per this Act. This brings to the forefront the importance of understanding that researchers should not simply share data without considering legal and ethical ramifications.

3.4.2 A teaching resource for science advancement

Data-sharing benefits science education, especially for faculty undergraduate and graduate students (Whitlock 2011: 63). It is increasingly tricky for some scientists, mainly students, to obtain data for their studies and research. As Whitlock (2011: 63) points out, making research data transparent is critical so that students and other researchers struggling to attain crucial review findings through informal routes can reuse it. For example, it may be difficult for students to obtain authorization to do an investigation in some organizations that do not recognize the significance of the research. As a result, these students only alternative is to use previously compiled data in their research work.

Peer review is another crucial concept for science advancement. Peer review is essential for confirming the validity of scientific research in academia (Wolfram *et al*. 2020:1033). Researchers have begun to doubt the validity of the reviewing process due to worries about the lack of transparency in the process (Besançon *et al.* 2020: 3; Zong *et al.* 2020: 608). The idea of open peer review (OPR), which aims to increase the credibility of peer review, was offered as a crucial innovation within the open scientific movement in response to these worries (Ross-Hellauer, 2017: 2; Wolfram et al., 2019).

Identifying reviewers or authors, publishing review reports, encouraging expanded involvement of various stakeholders, and creating channels for communication between authors and reviewers are just a few of the seven features of the review process known as OPR.

The term "open peer review" (OPR) refers to a review procedure that incorporates any number of the following seven features: disclosing reviewers' or authors' identities, publishing review reports, promoting more significant involvement of various stakeholders, opening channels of communication between authors and reviewers, publishing pre-review manuscripts, sharing final-version comments, and promoting an open platform where reviews are carried out by numerous institutional bodies rather than just one (Wei, Zhao, Ni *et al.* 2023: 2763).

3.4.3 Verification and reproducibility

The publication of research data helps the faculty preserve the accuracy of the research, allowing others to verify or recreate discoveries, boosting transparency, and improving research credibility. According to Cox and Pinfield (2014: 300), if scholars know that other researchers may re-analyze their data, they are more likely to produce better science since open access to research data fosters deeper examination. Sardanelli *et al*. (2018: 2330) concur that research data-sharing can lead to the discovery of new findings. New findings can be revealed based on assumptions not discovered by the original research team, and existing data can convey new insights not previously examined in the novel article (s). Scholars may also be interested in analyzing research datasets from several sources to improve accuracy, i.e., doing replicated assessments through multiple databases for existing concepts or novel ideas (Sardanelli *et al*. 2018: 2330). Another possible benefit of data-sharing, according to Qi, Deng, and Guo (2017: 500), is that it reduces the publishing of erroneous studies, especially when the data is purposely misrepresented.

3.4.4 Reduced costs and time in research conduct

A faculty research repository can save costs and time associated with conducting research. Several studies have shown that conducting research is not inexpensive (Qi *et al*. 2017: 500; Sardanelli *et al.* 2018: 2329). As a result, it makes financial sense to reuse data generated and provided by other researchers. According to Sardanelli *et al.* (2018: 2330), research data exchange could reduce the time and expenses of performing clinical research by avoiding trial duplication (Rowhani-Farid & Barnett 2016: 5). Costs such as securing patient insurance coverage, purchasing materials, and paying the salaries of data-gathering employees can all be avoided.

Furthermore, the new results could be available years before those obtained through a fresh clinical investigation using a current central repository. There is a need for the research community to develop solid data-sharing practices. Such solutions could involve the creation of online repositories where research data can be stored for later use. The risk is that researchers will lose resources such as money and time if research data is unavailable by collecting duplicate data (Chawinga 2019: 103).

3.4.5 Improving science

Science knowledge cannot be improved in a void. It must be built on earlier information, and faculties are the foundation of knowledge generation. If secondary data is well maintained, most researchers are encouraged to use it. Well-managed data is research data that is error-free, has sufficient documentation, and is described in detail (Tenopir *et al*. 2012: 70; Yoon 2015: 174; Yoon & Schultz 2017: 925). Reusers of data will also have confidence in data generated by primary researchers with expertise in data management (Yoon, 2015: 147). Understanding that the long-term availability of the data is essential for validating scientific advancement and building an infrastructure to handle future studies, research data should be carefully preserved. Researchers (Shakeri 2013: 71; Bond-Lamberty 2018: 1441; Dai *et al*. 2018: 131) concur with Tenopir *et al*. (2012: 71) that data-sharing aids in "expanding research from earlier results" and serves as a verification tool. Prior research has already demonstrated that the failure to share research data hampered fresh innovation and advancements.

The scholarly literature is still dominated by disputes regarding the benefits of research data, and researchers have taken a keen interest in data-sharing. The advantages of data-sharing were discussed throughout the literature, yet many researchers remain hesitant to share (Chiware 2020: 383). Chiware (2020: 383) contends that researchers are more intensely immersed in day-to-day scientific research and devote less attention to the benefits of sharing, reuse, and long-term accessibility to their data. Hence, the faculty research repository concept, which allows content to be curated in a single location, widening its appeal to a vast audience of study and instruction within the faculty, has proven helpful for the earlier adopters.

3.5 Research data repositories, design, and development

This section thoroughly looked at the state of data repositories that have been uncovered in research. The study offered insight into each data repository's functional features and tradeoffs of crucial design elements. According to Ng'eno (2018: 57), another method that could help preserve research data from HEIs is using data repositories, ensuring that researchers, scholars, and other stakeholders have access to, preserve, and disseminate research information. Data repositories can be found on Re3data.org, a global register of research data repositories from various academic fields. It provides researchers, funding bodies, publishers, and scholarly institutions with repositories for storing and retrieving data sets. Re3data.org encourages a sharing culture, enhanced access to research data, and greater visibility of research data.

3.5.1 Domain-specific research data repositories

These are specialized repositories that house research data from a particular discipline. Dryad for biology, the Inter-university Consortium for Political and Social Research (ICPSR) for economics, and the Marine Geoscience Data System (MGDS) for environmental sciences are just a few examples of RDRs. Domain-specific repositories are characterized by the ability to use specialized metadata for that field to improve, among other things, interrogation eloquence, indexing strategies, retrieval competence, and fine-grained pursuit outcome categorization and sifting (Borgman 2015: 125). One consequence of this intricacy is that it may be challenging or intimidating to the user, mainly when the individual is unfamiliar with the jargon used in the field of study. This could be a problem, considering that open data seeks to make scientific datasets widely accessible and reusable, including for the wider population. However, studies have indicated that once scholars are familiar with discipline repositories, they prefer to use them over other repositories (Borgman 2015: 125).

3.5.2 Institutional repositories

Higher education institutions offer repositories for the exclusive use of their scientific community, e.g., DUT has DUT Open Scholar. Institutional repositories are more often used to store university-produced creative and academic research outputs like books, patents, reports, conference presentations, research papers, master's theses, and doctoral theses than to store research datasets. According to Bugaje (2019: 57), institutional repositories are provided to participating institutions or their affiliated communities to preserve and disseminate digital assets developed by them. In most cases, institutional repositories only give fundamental and straightforward searching, sorting, and filtering capabilities. All the data in the repository is developed to enhance access and allocate permanent links. However, the drawbacks are that doing so may not yield any significance, given that a vast amount of data is not preserved in a single institutional repository. When available, institutional repositories' advanced search functions frequently provide options that are either more particular to text-based items or just slightly pertinent to datasets (Bugaje 2019: 57). Several institutions' libraries have an important function in developing RDM services and solutions. However, they are hesitant to do so since policies governing the management of research outputs, such as datasets, are still unclear (Chiware 2020: 383).

3.5.3 Publisher-based research data repositories

As noted by Bugaje (2019: 59), journal publishers provide publisher-based repositories, some of which perform a review process on research findings and then publish them as customary scientific outputs, popularly referred to as "data papers." Publisher-based repositories are primarily designed to link research data to the publications they support. Nevertheless, because journals publish on specialized areas or subjects, their repositories may share some of the benefits of domain-specific repositories. This service is not the most popular yet, but it is rising. Publishers who publish research datasets also produce journals, which could be a drawback. Their repositories typically house both types of resources and tend to adapt their differences by encapsulating research data criteria to those of published articles. Bugaje (2019: 59) argues that the search field options in such repositories are more helpful for discovering research publications than data.

3.5.4 Data repositories for commercial and general-purpose research

Since a general-purpose data repository is subject-agnostic and contains data from various fields, it stores multidisciplinary data from specialist disciplines lacking specific repositories (Borgman 2015: 125). General-purpose and commercial data repositories are commonly used systems with an extensive user base. General repositories are ideal for storing data since they typically include comprehensive features and institutional backing and are indexed by major search engines such as Google and Bing. The disadvantage of general repositories is that people may have more difficulty finding work because there is so much of it. According to Bugaje (2019: 60), general-purpose repositories, like many other types of repositories, give tools for browsing by subject, but they rarely provide further filtering choices for reducing search results.

A study by Zibani *et al*. (2021: 240) provided a thorough analysis of faculty research repositories at HEIs to understand their history, drivers, benefits, and adoption. The study evaluated the literature on research repositories utilized in HEIs and the justification for their development. Some factors were providing accessible, discoverable data, ensuring more visibility to academic output, encouraging faculty learning activities, and giving faculty and students access to research data members to share and reuse. According to Sweeper and Ramsden (2020: 11), developing a new research strategy and encouraging research principles at an institution requires a research repository. As institutions strive to become research leaders, promoting and presenting research outputs to a worldwide audience becomes more important. The research uncovered many repository systems within a HEI for various objectives.

The following are examples of repository platforms:

Non-commercial and commercial subject-based repositories are created by societies or communities (single and associated). Scholars intrinsically cherish the content, and this type of repository is distinguished by unprompted self-archiving. This type of repository enables the early dissemination of research conclusions and viewpoints in the form of draft publications. Pre-prints provide writers with benefits like the capacity to assert primacy, assess the significance of an idea or finding before submission, improve a journal work before submission, obtain recognition, draw attention from around the world, and more (Armbruster and Romary 2010: 62).

Research funding mandates inform research repositories and frequently make research data accessible and available to encourage reliability and accountability in the research process, enhancing corroborations and confirming study findings through repetitive data analysis to handle various research issues and assist breakthroughs (Elsevier 2019: 1; Ülikool 2020: 3).

National repository systems resemble networked systems designed to track scholarship and assist academic growth in higher education by collecting research output from researchers.

Institutional repositories include teaching and learning resources and the institution's varied outputs. The current research information system has features in some institutional repositories that are helpful for researcher profiling, grants, and publication activities (Ülikool 2020: 3).

According to the findings, faculty were adopting the idea of faculty research repositories for many reasons (Zibani *et al*. 2021: 240). One of the factors mentioned was the creation of open and reusable knowledge bases for systematic data collection. Other factors included uniformity, usability, simple administration and documentation of scientific outputs, learning projects, scientific data, and gateways for collaboration with peers in related fields. The review discovered that the concept of a faculty repository was well received. First-time adopters stated that it was completely utilized because it was believed to be appropriate to the demands of the faculty with essential data that addressed a range of interests and skill levels. The study also pointed out that some regions, including Africa, had not fully embraced the idea of faculty research repositories. One of the most critical explanations in this research study was the repository's potential to increase global awareness of its collection through simple access, discoverability, and sharing. The study also noted that although repositories showcase rare, unusual, and marginalized collections that are beneficial to faculty, to sustain repositories' value and relevance to faculty members and the specific disciplines they serve, it takes commitment and investment to build repository operations and procedures (Zibani *et al.* 2021: 240). The long-term survival of faculty repositories would depend on the capacity of academic research administrators to ultimately take ownership of these platforms and act as stewards of acquired knowledge. To claim ownership, one had to embrace and contribute to the repository as a resource platform instead of an operating system.

3.6 Criteria for selecting a research data repository

As governments, funders, and research institutions foster increased transparency and preservation of the information that underpins research findings, repositories are becoming more critical. According to Sansone *et al*. (2020: 3), thousands of research data repositories range from general to discipline-specific. Most of these platforms are for datasets, although a few are for software, pre-prints, and algorithms. Sansone *et al*. (2020: 3) also note that many repositories are starting to include the FAIR Principles in their policies and implement the necessary technical changes to enable the data's Findability, Accessibility, Interoperability, and Re-usability. Simultaneously, publishers and journals adopt data policies to ensure that datasets and other digital assets linked with articles are deposited and made available through appropriate repositories. With thousands of alternatives, however, publishers' lists of deposition repositories are frequently varied, and as a result, the advice given to writers varies from journal to journal. This is due to a lack of standardized criteria for selecting data repositories and a lack of consensus on what constitutes a robust data repository. A few organizations, including the domain-agnostic CoreTrustSeal (CTS), the emerging TRUST principles, the domain-specific ELIXIR Core Data Repositories, and the funder-specific guidance from the USA National Institutes of Health (NIH), have proposed requisites for recommending and assessing the caliber of data archives (Sansone *et al.* 2020: 3). Despite being characterized differently, these various projects aim to specify criteria, some of which are identical.

3.6.1 Trusted data repository elements

A critical review and reflection on the essential elements in the prototype design for a data repository were dealt with. The section presented some ideal aspects suggested in the study by Van den Eynden (2013: 18) for choosing a repository to manage and exchange data.

3.6.1.1 Sponsorship and governance

Governance is critical to determining a repository's and its data's long-term health. Another technique to assess the repository's long-term viability is to look at its funding source. Government funding is determined by legislative expenditures, which are contingent on the political party in power and the state of the economy. However, repositories run by significant government bodies (such as the National Institutes of Health) are unlikely to shut down. Foundation-funded repositories are vulnerable to grant shortages or even termination. Organizations that receive financing from various foundations (such as the Center for Open Science) are more likely to survive.

3.6.1.2 Preservation

Understanding the operational and monetary resources a repository utilizes for data backup is an essential evaluation criterion. A repository is more likely to exercise great caution in carrying out this vital work if it clarifies how it provides durability, geographic spread, and the payment structure for such services. Examining the organization's strategy for data extraction by proprietors if the repository is shut down or the organization runs out of funding is also crucial.

3.6.1.3 Pricing

Others demand a fee depending on how big the files are in one's dataset, while some repositories are free and open to the public. A free alternative is typically offered for datasets of a particular size, with further tiers accessible over specific storage benchmarks. The potential size of the dataset and whether it will result in increased storage costs must be considered. Some providers offer a free option for public storage and a cost for private storage.

3.6.1.4 User support services

Frequently asked questions (FAQs), repository documentation, instructional films, and employees' access to answer queries are all support services a repository provides. Depositing and accessing data can be difficult in a repository with inadequate user support. This criterion is significant for projects involving multiple users that want to store and access data in a repository, which may call for coordination and in-depth repository knowledge.

3.6.1.5 File Formats

Although most repositories recognize various file types, there is no assurance that the research will be accepted. To ensure that research data is appropriately formatted, see the FAQ and other documentation for the repository. Data portability is also crucial. Commercial software or other specialized tools are occasionally used to collect and record research data. If the research utilizes specialist programs, making that dataset in a standard and portable format would benefit other researchers. Comma Separated Value text files (CSVs), XML files, PDF files, and raw text are examples of formats.

3.6.1.6 Space

A repository's usefulness may depend on the size of the individual files in the study dataset and the overall amount of data. For instance, a repository might restrict the maximum individual file sizes that can be uploaded, but it might not restrict the total quantity of data that can be deposited. If data includes large image files larger than 1GB, they may exceed a repository's file upload limit. Due to network limitations, cooperation with university IT may be required for some repositories that embrace large quantities of files.

3.6.1.7 Versioning of data

Typically, research efforts create many versions of data. If the research project requires numerous contributors, versioning or monitoring file changes is critical. Furthermore, future researchers must be able to cite and access the factual information that supports their research if one's published data is ever updated.

3.6.1.8 Structure of the original directory

In recent years, several scientific discoveries have been called into doubt or completely disproven due to their inability to be replicated. Due to the "reproducibility dilemma," many journal publishers and foundations require scientists to make replication data available so that third-party researchers can confirm findings. The repository must allow a researcher to retain directory structures when entering their data precisely to call, analyze, and produce outputs to duplicate computational or statistical conclusions (e.g., tables).

3.6.1.9 Metadata

Metadata is organized data that can help researchers find, interpret, and reuse data more efficiently. The identity of the research project's lead investigator, the timeline for data collection, and the license connected with the data are all examples of dataset metadata. When choosing a repository, ensure it adheres to widely accepted metadata standards, especially in the researcher's discipline. Whilst using metadata standards whenever feasible is desirable, check if the repository allows the researcher to alter the metadata fields to meet their specific needs.

3.6.1.10 Unique identifiers

Making research more visible and citable is one of the foremost aims of contributing data to a repository. By giving the content a citable reference and enabling other scientists to obtain it years after one has published it, a DOI (digital object identifier) or handle increases the visibility of research data.

3.6.1.11 Permissions and roles

If numerous contributors work with the data, consider data repositories that enable researchers to set up unique roles and permissions. Some team members could need administrator rights, but others might only need submission or read-only access. Verify the data repository's review process for new or existing content revisions.

3.6.1.12 Access

There are times when restricting access to research data is necessary or desirable. Research data, for instance, can include private data, patient information, or sensitive material. Moreover, the publisher might have restricted the data, or the data might still be in the development stage and not be ready to be shared.

3.6.1.13 Licensing and rights

A repository's license and rights regulations should explicitly state service and FAQs. Data visibility may be restricted by stringent licensing regulations in repositories or require users to relinquish rights they want to keep. Some may offer more approachable licenses, like Creative Commons licensing, but only for a price. The terms of service should also spell out a user's rights if a repository goes out of business or raises its prices.

3.7 Metadata standards and interoperability

Metadata is often used for information retrieval searching to allow users to locate available data efficiently and record references for verifiable sources. It is an essential component within the repository structure; thus, it was crucial to review. As observed by Van den Eynden (2013: 14), metadata is a significant proportion of the fundamental pre-defined and organized data documentation that, in the setting of data processing, designates the originator, intention, creator, period citation, location, and conditions of use of a data collection. Furthermore, Shakeri (2013:10) posits that metadata can be created using the dialect of the data sources. For data to be reused, metadata is necessary since it communicates its legitimacy, worth, and relevance to researchers. Metadata contains detailed information about the original data source (e.g., organism, research lab specimen); procedures used to construct the data (such as experimental design and environmental elements); and details on the citation of specific data (Wittig *et al.* 2017: 229). As stated by Shakeri (2013: 10), metadata for online data collections and discovery access points are frequently structured following international standards or schemes such as Dublin Core, ISO 19115 for geographic information, Data Documentation Initiative (DDI), Metadata Encoding and Transmission Standard (METS), and

General International Standard Archival Description (ISAD(G) (Shakeri, 2013:10). Using standardized records in eXtensible Mark-up Language (XML) combines critical rich and organized data material as a result of combining all data information into a single document. Internet browsers can showcase metadata; harvest and analysis engines can utilize it; and field searches can be done. It is possible to exchange catalogs from many sources and use dynamic searching features. Furthermore, The Open Archives Initiative Metadata Harvesting Protocol (OAI-PMH) allows metadata collection for data exchange.

Numerous efforts to define the meaning of interoperability have resulted in the following examples:

> Abbott (2009: 27) defines interoperability as the efficient and consistent sharing and use of information across numerous organizations, systems, and platforms. "The ability of numerous systems with diverse hardware and software platforms, data structures, and interfaces to exchange data with little loss of content and functionality" (NISO 2004: 17).
>
> "Interoperability is the ability of two or more systems or components to share information and use the information exchanged without particular effort on either system" (CCDA, 2000:31).
>
> "Interoperability" is defined as "the compatibility of two or more systems such that they can interchange information and data and use the exchanged information and data without particular manipulation" (Taylor 2004: 369).

As previously stated, metadata is critical to data access, sharing, and reuse. Regarding data sets, interoperability is frequently mentioned in the literature as a significant impediment to data sharing and reuse due to a lack of metadata and formatting standards. (Chawinga 2019: 41; Woolfrey 2009: 75). According to Brown, Bruce, and Kernohan (2015: 85), there are two categories of existing methods and solutions for data interoperability: approaches that are arrangement and neutral third party. Interoperability issues are frequently complicated and require myriad issues to be remedied. Even in this instance, solutions are also accepted or mediated. In some circumstances, agreement-based and mediator-based approaches merge. Agreement-based techniques entail agreeing based on values that allow for a constrained measure of uniformity among diverse organizations (Brown *et al*. 2015: 85). As Shakeri (2013:

79) noted, using standards is part of interoperability. Although it is simple to demonstrate the value of standards, most organizations embrace them only after carefully weighing the benefits and drawbacks. Standards are frequently complex combinations of features that reflect the interests of numerous stakeholders. As a result, they are not always simple to implement (Shakeri, 2013: 79). Additionally, they violate the adopted organization's freedom. This implies that businesses must agree to change their behavior to improve interoperability. Large-scale measures based on agreements and initiatives to advance data interoperability include Linked Data (Heath & Bizer 2011: 11), INSPIRE (European Parliament, Council, 2007), SDMX (SDMX Initiative n.d.), and OAI-ORE (Lagoze & Van de Sompel 2008: 35). These critical factors are crucial when making decisions on the choice of the repository for a faculty to enable search, discoverability, and access to content.

3.8 Repository system designs and development

This section reviewed components critical in designing the faculty research repository prototype conducive to a UoT, which was presented and discussed in the research study's succeeding chapters. According to Satzinger *et al*. (2016: 5), system design refers to acts that make it possible to describe how the final system will be implemented fully. The main elements of RDM platforms consist of a user experience, a recovery method, and a repository or file storage structure (Bugaje 2019: 28). With the procedures and functionalities of the modules fully or partially integrated, the development and design process can begin at any RDM building scale. Some RDM platforms, for example, merely provide a user experience for searching and discovering information, depending on third-party retrieval algorithms for inquiry optimization and external repositories for datasets (Bugaje (2019: 28). As asserted by Witt (2008: 195), the points of connectivity and metadata required for browsing and surfing data repositories are described by research data discovery systems, along with methods for supporting external users and user agents like search engines in discovering data. In contrast, in other search terms, RDM platforms support inquiry and exploration execution architectures but depend on different databases to provide the essential datasets. There has not been much research on the design and development of RDM platforms, making it difficult to compare them to the variables and methods used to arrive at the design and development of the systems. Even though no precise standard design method for any of the RDM systems discussed in the previous section has been found, this study will review the user-centered design approach, prototype design, and development because they are crucial components of the research.

3.8.1 User-Centered Design

Bugaje (2019: 31) stated that rules and best practices for designing RDM systems have yet to be created. There is a risk of overemphasizing the platform, the functionalities, and the abilities it must offer while experimenting with novel concepts and techniques rather than the system users and their needs. When creating a system, the user-centered design method considers the needs and situations of the people who may use it. Various methods define the user-centered design depending on the developed product's intended use and target market. Ames (2001: 138) defines it as a procedure that includes examination, design, user authentication, development, and implementation. Indeed, user involvement appears to be seen as a prerogative of users rather than a requirement in user-centered design. According to the International Organization for Standardization ISO/IEC (1999: 11), the critical component is user involvement throughout the product design. Usability is determined by the International Organization for Standardization (ISO) "as the degree to which specific users can utilize a product to accomplish particular goals in a satisfying, effective, and efficient manner specific context of use." This definition is used to describe how a user-centered design process is conducted. According to Bowler *et al.* (2011: 730), user engagement is a hallmark of user-centered design, intending not simply to produce something that functions but only for the target audience. In addition to the inherent benefits of the approach, one of the main justifications and arguments for using the user-centered method in creating RDM platforms is that research data repositories are end-user systems designed with the consumer in mind. In such a scenario, user contentment becomes a critical pro or con signal of whether the system will succeed or fail (Bowler *et al*. 2011: 730). The following are the tenets of user-centered design that Satzinger *et al*. (2016: 221) suggest serving as the center of the RDM platform:

- Prioritize the user and their work early on and throughout the entire project;
- Verify the usability of all designs; and
- Adaptive development is used.

Cimino *et al*. (2014: 18) asserted that the NIH Biomedical Translational Research Information System (BTRIS) is one of the repository platforms whose development process adhered to "good software development practices. It prioritized four fundamental requirements in its development process, namely:

- The capacity to handle any data that could be encountered;
- A database design that is prepared for the kinds of queries that are most likely to be made;
- Adopting a regulated vocabulary would include the precise phrases in the data and the slightly elevated constructs that users will probably bring up in their inquiries; and
- NIH researchers could conduct their searches with the help of a user interface (Cimino *et al.* 2014: 18).

3.8.2 Prototype design and development

Prototyping is a core engineering design activity and an important research issue (Gero & Lindemann 2005: 17). Wall *et al*. (1992: 170) consider phototyping among the most crucial steps in developing new products. However, the various definitions of "prototype" listed by Jensen, Özkil, and Mortensen (2016: 826) show that advanced engineering prototypes still have considerable data and investigations to be done. According to Erichsen *et al*. (2021: 67), prototyping mainly produces tangible results through prototypes and intangible results in the type of expertise, abilities, and perspectives of the design team. Consequently, Prototypes are considered the outcome of development and testing. This contrasts with the definitions provided by Jensen *et al*. (2016: 826), who argue that prototyping is merely the process of creating prototypes. According to Jensen *et al*. (2016: 826), a prototype begins with moderate approaches such as utilization, storyboarding, and paper prototyping. They assist in the low-level design of the product; they explain specific use cases and user activities in general; and they outline the features that the user should be able to access on the platform. Paper prototypes are produced to understand user requirements better and evaluate their effectiveness. Following that, the complexity of the prototypes is raised, starting with the mock-up paper prototyping method and concluding with a high-quality functional interactive product (Wall *et al.* 1992: 170). The usage of mock-ups enables the researcher to communicate the design and explain functioning while allowing the user to test the structure early on and make adjustments as necessary.

3.8.3 Objectives for prototype design

According to Camburn *et al*. (2017: 3), design prototyping has numerous potential goals. This section delves into the common goals of prototyping. The following are the most frequently reported aims (Camburn *et al*. 2017: 3):

3.8.3.1 Refinement

Prototyping involves the process of continually enhancing a design, known as refinement. Several advantages of prototyping are related to improving the conceptual design. Prototyping is used to verify necessities, identify essential design issues and minimize errors (Viswanathan 2012: 79); identify design adjustments that will improve performance (Viswanathan & Linsey 2011: 5; Viswanathan & Linsey 2012: 37); optimize design features through consecutive analysis and parameter handling (Anderl, Mecke & Klug 2007: 505); and solo or many tests that imitate actual use can be used to improve the design (Otto & Wood 2001: 18).

3.8.3.2 Communication

Interaction between consumers and design team members involves exchanging information regarding the layout and possible applications. According to design research, prototypes are vital for exploring and improving design usability (Gordon & Bieman 1995: 90; Barbieri *et al*. 2013: 311). The reflection of actual user collaboration with one another and design decisions can be made using prototypes (Kurvinen, Koskinen & Battarbee 2008: 48). Furthermore, prototypes are useful for sharing ideas between members of the design team (Buchenau & Suri 2000: 425). This objective includes usability testing to look into human factors in design.

3.8.3.3 Exploration

The practice of finding out fresh conceptual designs is known as exploration. Divergence and convergence are high-level design processes that might be coupled with prototyping. Divergence means acquiring knowledge and developing a variety of new conceptions. In contrast, integration suggests narrowing the set of sophisticated ideas (Lennings *et al*. 2000: 7). According to industry research, professional designers commonly employ physical prototypes to assist in the idea-generating process (Hess & Summers 2013: 249).

3.8.3.4 Active learning

Active learning is acquiring novel concepts about the design space or pertinent phenomena. Active learning in this setting refers to expanding designers' mental or analytical models of supernatural encounters; practical prototyping exercises help with the acquisition of implicit knowledge about multifaceted occurrences (Telenko *et al*. 2016: 2). Physical models enhance design education by showcasing adaptability across a variety of domains (Green & Smrcek 2006: 192). From the designer's psychological perspective, prototypes are crucial for reassessing failure as a chance to learn and advance a sense of progress (Gerber & Carroll 2012: 65). Physical model development can aid in identifying contrasts between a notion and actual behavior.

Numerous research has been conducted to determine incorporating prototyping into the design process. This section summarizes the significant results of prototyping and the design process.

3.8.3.5 Testing

Validation is the main feature that separates a prototype from a proposed design. However, completing each test incurs costs or effort (Dahan & Mendelson 2001: 103). One of the primary goals of prototype design methodologies is to gather enough knowledge to move forward with product development while spending as little time and money as possible. As a result, each prototype experiment should answer a particular question (Otto & Wood 2001: 18).

3.8.3.6 Timing

According to empirical and industrial studies, early prototyping is crucial for innovation (Rothenberg 1990: 403; Drezner 1992: 13; Jang & Schunn 2012: 11), particularly during the first 30 percent of a design project (Otto & Wood 2001: 19). In contrast, late prototyping is associated with failed endeavors (Jang & Schunn 2012: 11). Appropriate prototyping timing should be considered strategically (Otto & Wood 2001: 19), but this selection may be influenced by complexities (Schrage 1993: 56). It is vital to realize that devoting more time to prototypes does not equate to more success (Yang 2005: 650). It is thus critical to begin positioning efforts early in the overall schedule.

3.8.3.7 Ideation

For proactive concept exploration, prototypes are vital because they promote spontaneous learning and exploration (Gill, Sanders & Shim 2011: 27; Hess & Summers 2013: 249). Early prototyping frequently results in new design concepts (Yang 2005: 650). Outdated development techniques can be flipped to employ prototyping for creativity, such as molding a clay object and scanning it. Prototype-generated ideas are more likely to be functional (Viswanathan & Linsey 2012: 25).

3.8.3.8 Fixation

When there is an exemplar response, prototyping to ideate rather than pure sketching results in fewer focused solutions (Youmans 2011: 125; Schubert *et al*. 2012: 75). It is vital to note that, compared to sketching alone, a slow production process may cause fixation, while a fast one will not (Viswanathan & Linsey 2013: 122). Prototypes may be used more frequently to communicate inside domain analogies, while sketches may communicate between domain analogies (Christensen & Schunn 2007: 31).

3.8.3.9 Feedback

Designers might feel uncomfortable showcasing prototypes to supervisors or managers. Moreover, they might feel more comfortable doing so to their peers (Schrage 1993: 56). Stakeholders in several prototyping cultures demand high-fidelity prototypes. If they receive low-fidelity prototypes, they may dismiss a venture (Schrage 1993: 56). Feedback should be appropriately timed and structured to avoid disrupting perception and reward cycles (Schrage 1993: 56). It is also worth noting that receiving comments on a design concept may encourage fixation on that particular design concept (Kershaw, Hölttä-Otto & Lee 2011: 81). This can be reduced by displaying various designs (Dow *et al*. 2009: 5).

3.8.3.10 Usability

During testing, it is critical to record the customer's (end-users) voice (Christensen & Schunn 2007: 31). The level of interactivity is one aspect of prototyping fidelity (Dow *et al*. 2009: 5). The design team can determine how well a prototype captures anticipated user interactions, which will, in turn, define the scope of testing (Kershaw *et al*. 2011: 81).

3.8.3.11 Fidelity

An experiment has shown that better quality representations result in more accurate interpretations by third-party design reviewers (Hess & Summers 2013: 11). This results in a trade-off with the observation that faster prototyping leads to reduced fixation. As a result, the right level of faithfulness must be chosen based on the goal of a particular effort.

3.9 Critical challenges in RDM

As the focus on developing data repositories continues to grow, research data management challenges are dealt with by research institutes and universities as part of intricate data management activities. According to Wilkinson *et al*. (2016: 9), the current digital environment surrounding scientific data publication prohibits the maximum advantage of realizing research efforts. The scope of this study was to design a prototype for a faculty research repository conducive to a UoT that would help overcome at least some of the problems associated with RDM systems. Therefore, it was necessary to be aware of and comprehend these difficulties to the greatest extent feasible. The challenges could be technical, socio-cultural, or ethical (Curdt & Hoffmeister 2015: 500; Bugaje 2019: 37; Chawinga 2019: 43). This section discusses a few of these difficulties.

3.9.1 Inadequate Metadata

This was a significant impediment to offering rich access and finding capabilities for research data (Kouper *et al*. 2013: 43; Borgman 2015: 128; Bugaje 2019: 37). When data are inadequately defined, prospective reusers cannot comprehend the context or quality of the information or how they were generated to any practical level, rendering re-purposing exceedingly difficult (Koltay 2017: 5). Mainly, research data owners or researchers carry the majority of the responsibility for metadata labeling and documentation, whilst several research findings have shown that scholars tend to dedicate little time to this action (Carlson *et al*. 2011: 640; Wallis *et al*. 2013: 12; Chowdhury *et al*. 2014: 27; Borgman 2015: 128).

3.9.2 Inadequate RDM Skills

Bugaje (2019: 38) observed that the knowledge of research and scholarship does not inherently entail an understanding of data management. Similarly, several studies on academic researchers have indicated a significant skills gap between what is required of researchers in their function as data creators and what their general level of expertise allows them to accomplish (Cox & Pinfield 2014: 308; Davidson *et al*. 2014: 299; Borgman 2015: 128). Data discovery strongly relies on good metadata and data authors, even though the significant sources of contextual metadata and other supplementary information about data are not often proficient in data management or informed about its intricacies (Willis *et al*. 2013: 12; Borgman 2015: 128; Bugaje 2019: 39).

3.9.3 Inadequate RDM infrastructure support

Bugaje and Chowdhury (2017: 54) observed that the current RDM infrastructures are insufficient for supporting researchers in meaningfully conveying data. The current inadequacy of RDM systems is because they are still an improvised modification of text or thread information management rather than solutions created with the specificity, complexity, and needs of research datasets in mind. Moreover, most university assistance for research data preservation is limited to implementing high disk storage and backup solutions. Shared folders are sometimes the only means of collaboration between researchers (Chawinga 2019: 48). Chawinga (2019: 48) notes that a small percentage of UK universities possess potential storage that is adequate for a while. However, they encounter technical and organizational hurdles in delivering an effective hybrid for storage because their equipment is dispersed over many faculties or sites. According to a study conducted by Chen and Wu (2017: 346) in China, the most common difficulty faced by researchers was unreliable storage infrastructure, which resulted in frequent data loss. The research commissioned by the European Union revealed that the most significant challenges to digital preservation in Europe are a lack of long-term hardware, software, and computer environment support (Chawinga 2019: 49). The infrastructure challenges have been observed by Ng'eno (2018: 77), citing weak data infrastructure as one of the significant hurdles facing RDM efforts in Africa, notably Kenya. Poor data infrastructure results from ongoing funding constraints, according to an international qualitative study covering Uganda and Tanzania (Anane-Sarpong *et al*. 2017: 9).

3.9.4 Significant time constraints for researchers

Data management procedures take time, and researchers have limited time. In many cases, the advantages or rewards are not immediately apparent. As a result, academics prefer different scholarly endeavors, including producing scientific papers with more demonstrable outcomes or having better esteem in the academic community, like paper citations and h-index (Borgman

2015: 128; Chawinga 2019: 51). According to Borgman (2015: 128), the majority of researchers believe that the resources and time used for data management were wasted.

3.9.5 Policies enacted by institutions

According to the literature, institutional rules and guidelines significantly encourage or discourage data sharing (Walters & Skinner 2011: 31; Chawinga 2019: 51). Under this setting, institutional policies are standards produced at the institutional level to define which data sets are suitable for applying organizational resources to curate (Walters & Skinner 2011: 31). This shows that not a repository's curating capabilities may benefit all research data. Some entities may prioritize specific data sets. For instance, health organizations like medical and dentistry colleges can concentrate on archiving data from medical research while ignoring data from research on information-seeking behavior (Chawinga 2019: 51).

Promoting the accepted citation style and enticing researchers to reuse and cite other people's data sets are crucial aspects of research data management (Brown *et al*. 2015: 85). Elsayed and Saleh (2018: 282) conducted an online survey with 337 participants and found that the use of data citations is already actively supporting researchers in some Arab universities, such as those in Egypt, Jordan, and Saudi Arabia, to share data. The study found that 64.4% of researchers made their data available for citation so that they would remain well-known in their respective scientific disciplines. Given that when researchers are aware that their work will be referenced, they are more easily persuaded to submit their data to a repository, as stated by Patterton *et al.* (2018: 22), the findings reported by Elsayed and Saleh (2018: 282) are not astounding. Therefore, the results reported by Elsayed and Saleh (2018) are not unexpected. Universities are ranked using publication and citation counts by the highly regarded university ranking systems QS World University Ranking and Times Higher Education World University Rankings. The discussions above highlight the requirement for a central location to access and discover research data.

3.10 Appraisal of the Chapter

Chapter three provided a comprehensive literature review that delved into the core themes, objectives, and theoretical foundations of the research problem. It effectively synthesized existing knowledge by systematically analyzing pertinent studies related to research repositories within HEIs. The critical analysis commenced by assessing essential aspects of managing existing research, covering research data creation, evaluation, preservation, disposal, and storage facilities. Notably, it critically examined research data-sharing and reuse by researchers, shedding light on their utility.

One of the salient features of the chapter was its discussion on the fundamental components of RDM systems, emphasizing the significance of a user-centered design methodology, as well as prototype creation and development. Including these aspects enriched the understanding of the mechanisms involved in efficient RDM.

Furthermore, the literature review meticulously addressed challenges impeding the effective execution of research data management within research institutions. These challenges encompass inadequate metadata, limited RDM knowledge and experience among researchers, absence of legal and ethical guidelines due to rapid technological advancements, institutional RDM policy gaps, and time constraints for researchers. Importantly, it highlighted the early stage of research data management and the time-intensive nature of framework development, underscoring the need for continued efforts to overcome these challenges.

A critical contribution of the literature review was identifying a significant gap concerning limited faculty research repositories in South Africa. This observation highlighted an area for potential enhancement. It underscored the need for further research and development to bolster regional research repositories, ensuring robust data management and accessibility for academic and scientific advancement.

The research methodologies will be described in the next chapter.

CHAPTER FOUR: RESEARCH DESIGN

4.1 Introduction

The methodology for the study is described and justified in this chapter, based on the research topic and objectives given in Chapter One, and supported by the review of the literature and theoretical frameworks covered in Chapters Two and Three. According to Kallet (2004: 1229), research design describes the steps that will be taken to investigate a research problem as well as the justification for employing techniques to find, choose, process, and fully comprehend data to understand the problem better, enabling the reader to evaluate the study's overall validity and reliability. The section responds to two essential queries: How was the information gathered or created, and how was it interpreted? Byrne (2017: 1) mentioned that a research project could address new or current problems, improve recent work in the field, reinforce earlier work's results, tackle new or present challenges, validate preconceived notions, or develop novel approaches. To check the accuracy of the instruments, the research may rerun parts of earlier experiments or the complete project. The main objectives of the research include recording, discovering, interpreting, investigating, and advancing methodologies and systems for data improvement (Byrne 2017: 1).

The study's goal was to examine how research data are managed within the faculty environment of a university of technology to support researchers by creating a platform for a research data repository appropriate for a university of technology. The study employed an explanatory sequential mixed-method research design with a two-step process. Initially, a quantitative approach involved conducting an online survey for postgraduate students pursuing Master's and Doctoral degrees between 2015 and 2020. Subsequently, a qualitative phase was executed, involving a meta-analysis of research repositories across global HEIs and online interviews with postgraduate supervisors who hold key roles in postgraduate study and research administration. The explanatory sequential mixed-method study design consists of two successive phases: a quantitative phase devoted to gathering and analyzing numerical data, followed by a qualitative phase devoted to gathering and evaluating textual data (Ivankova, Creswell, & Stick 2006: 11). There are different approaches to connect and integrate the two components during the intermediate stage, with the quantitative phase typically taking primacy. The graphic below illustrates how the study's research technique was organized and modified from Creswell and Plano Clark's (2018: 109) mixed-methods design.

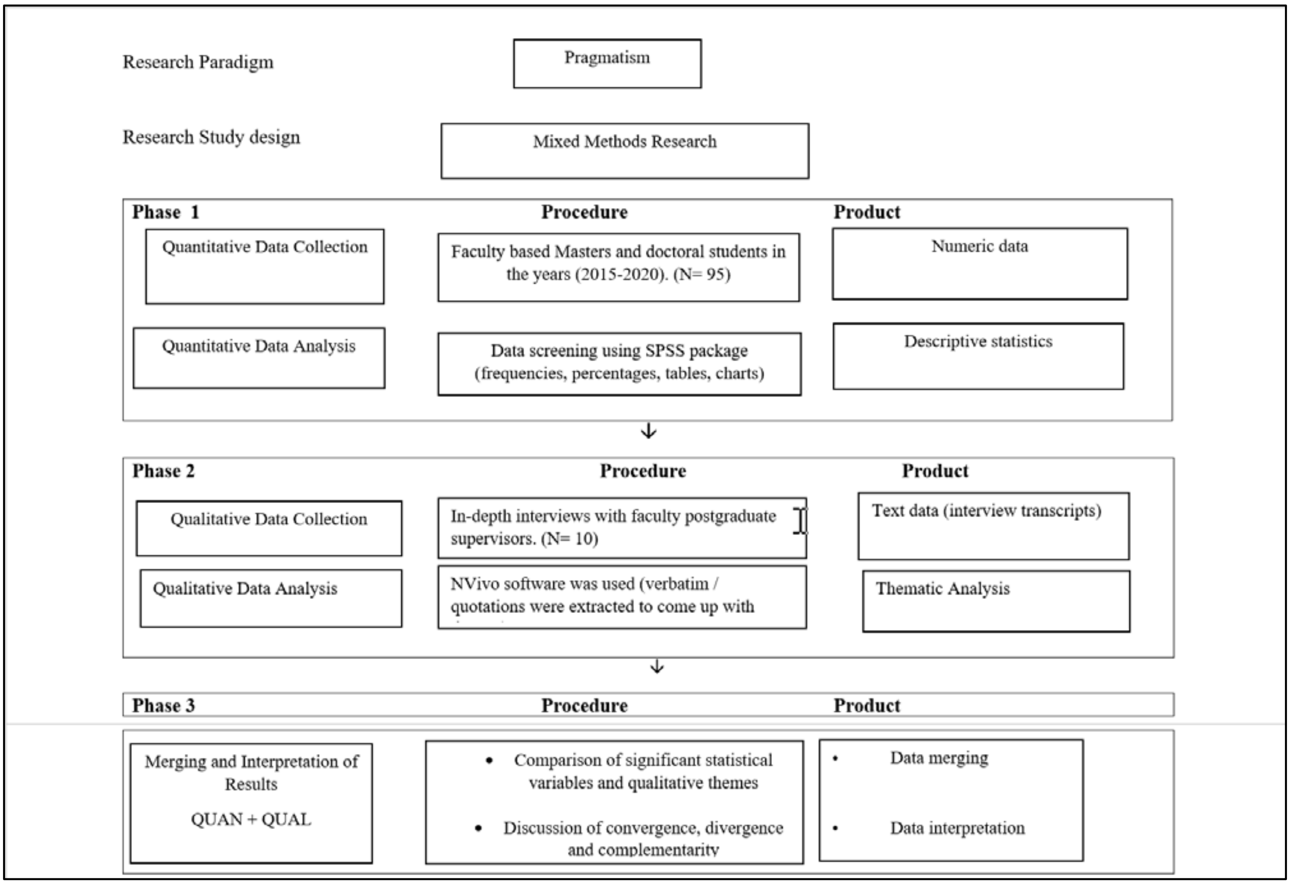

Figure 4.1. Researcher's adaptation of the Explanatory sequential mixed methods design.

Source: Creswell and Plano Clark (2018)

4.2 Philosophical World View

A research community's common preconceptions and views about the world are known as research paradigms. These preconceptions and views act as foundations for directing researchers' perceptions and behaviors (Mackenzie & Knipe 2006: 12). This worldview presents a perspective way of thinking, philosophical school, or shared values that influence research findings understood or interpreted. A research paradigm invariably reflects the researcher's views on the context in which they live and aspire to live (Lather 1986: 1230). It consists of abstract ideas and beliefs that affect how a researcher sees the world, interprets it, and behaves. It serves as the prism through which a researcher perceives the outside world. To determine the research methods and the data to be studied, the researcher analyzes the methodological components of their research project through a conceptual lens (Guba &

Lincoln 1994: 105; Denzin & Lincoln 2005: 29). The philosophical worldview of this research is pragmatism.

4.2.1 Pragmatism

Pragmatic philosophy aims to reconcile more recent techniques' naturalistic and unrestricted direction with the scientific method and structuralist focus of previous ways (Creswell 2018: 24; Creswell & Plano Clark 2018: 99). Due to having a conceptual basis in the philosophical insights of pragmatism throughout history, the pragmatism research paradigm is open to various methodologies. According to pragmatic thinking, researchers should use the methodology or philosophy most appropriate for the study question (Tashakkori and Teddlie 1998: 23). In research using mixed or multiple methods, the emphasis is on the study's results and hypotheses rather than the methodologies (Creswell & Plano Clark 2018: 99). Both formal and informal discourse is possible (Creswell & Plano Clark 2018: 99). A research challenge can be solved according to the pragmatism paradigm, which combines concepts, techniques, strategies, principles, or a mix of these. Pragmatism emphasizes what can be done or what functions more than the positivist principle of objective truth or reality.

The study applied the pragmatism paradigm since it concentrated on the research problem and employed various methods to learn more about it (Rossman & Wilson 1985: 630; Creswell 2009). The study's objective was to create a platform for a research data repository suitable for a university of technology by investigating how research data are maintained inside the faculty environment of a university of technology to help researchers. Both closed and open-ended survey questions from an online platform and interviews were used in this study to collect quantitative and qualitative data. Creswell (2018: 11) asserts that embracing a practical paradigm enables researchers to use various methods, viewpoints, and presumptions in multiple data collection and analysis types. The pragmatic paradigm has been applied in related and comparable investigations. For instance, Ng'eno (2018) used pragmatism in a study examining RDM at the Agricultural Research Institute in Kenya to provide initiatives to improve agricultural research findings' administration, dissemination, and repurposing. Similarly, Chawinga (2019) employed a practical strategy in his investigation into the administration of research data in Malawi's public institutions to create best practices for doing so in these and other research settings.

4.3 Strategy of Inquiry

According to Kirshenblatt-Gimblett (2006: 95), the research study examination strategy is the overarching approach chosen by an investigator to incorporate the numerous characteristics of the investigation systematically and coherently, assuring that it will successfully tackle the study's inherent problem. It is the groundwork for data gathering, evaluation, and assessment. A research design ought to be sure that the data collected allows the researcher to succinctly and successfully resolve the research dilemma. In social science research, gathering data relevant to the research inquiry entails evaluating the data required to verify an idea, consider a program, or accurately describe reality (Gall, Gall & Borg 2006: 181). Research designs include quantitative, qualitative, and mixed methodologies. A study design aims to verify that the information gathered allows the researcher to respond precisely to the original question. In other words, it is crucial to bear the following when conducting research: "What kind of evidence, given the research topic, is necessary to offer a convincing response" (Polit and Beck 2004: 120).

Moreover, Creswell (2018: 12) and Kothari (2004: 3) emphasized the relationship between various study designs and other philosophical suppositions and research procedures. For instance, a few examples of quantitative methods are post-positivist research designs that use descriptive, correlational, experimental, survey, and comparative quantitative methodologies. The social constructivism paradigm-related study designs also use qualitative approaches like ethnography, phenomenology, grounded theory, and case studies. Furthermore, the pragmatic paradigm that employs mixed approaches is paired with research designs such as concurrent parallel design, explanatory sequential design, exploratory sequential design, and embedded design. To compare or combine quantitative and qualitative results and draw more thorough and reliable conclusions, the researcher in this study used concurrent quantitative (web surveys) and qualitative (interviews) design methodologies. The strategy of inquiry employed is Explanatory Sequential Mixed Method Research. As mentioned in section 4.1, the execution of an explanatory sequential design involved two distinct steps. Initially, a quantitative strategy was implemented through an online survey targeting postgraduate students pursuing Master's and Doctoral degrees from 2015 to 2020. Following this, a qualitative phase ensued, encompassing a meta-analysis of research repositories across international HEIs and online interviews with postgraduate supervisors occupying crucial roles in postgraduate study and research administration.

4.3.1 Mixed Methods Research

Johnson and Onwuegbuzie (2004: 14) posit that the triangulation, or corroboration, of data from many sources, is made possible by mixed methods research, which integrates and evaluates statistical data with broader context-dependent insights. The goal of researchers who employ a hybrid method is to optimize the advantages and disadvantages of qualitative and quantitative techniques (Johnson & Onwuegbuzie 2004: 14). In a specific experiment; researchers may employ both quantitative and qualitative research methods, known as a mixed-methods approach to combine data collection and analysis techniques (Creswell 2018: 153; Johnson & Onwuegbuzie 2004: 14; Tashakkori &d Teddlie 2003). Researchers can now test and develop hypotheses by planning research studies that blend data gathering or data inquiry techniques from quantitative and qualitative research methods. Deductive and inductive reasoning can be used in the same research investigation. Investigators can utilize the mixed methods approach to research to produce a solitary study that responds to inquiries about the complexity of events from the standpoint of participants and the relationship between quantifiable aspects (Sale 2006; Johnson & Onwuegbuzie 2004). To investigate, forecast, find, describe, and understand the notion, enthusiasts of the mixed methods of investigation encourage doing "what works" within the guidelines of the study (Sale 2006; Johnson & Onwuegbuzie 2004; Creswell 2018;).

The mixed methods research, which is frequently associated with the pragmatic paradigm, was used in this investigation (Creswell 2018: 24; Creswell & Plano Clark 2018: 99). This approach was adopted because it bases knowledge claims on practical concerns and emphasizes the most efficient ways to deal with the research issue (Tashakkori and Teddlie 1998: 23). Both qualitative and quantitative designs have biases (Creswell 2018). The deliberate integration of quantitative and qualitative methodologies within a single study or across a series of studies distinguishes mixed methods research from other types of research. This amalgamation is hoped to arrive at a comprehensive grasp of a research problem. The idea of mixed methods research was created to solve the shortcomings in each research approach. To reinforce the insights received by the complementary method, the study used a mixed-method strategy that combined different methods to give more complete data, corroborate findings, and leverage results from one method (Creswell & Plano Clark, 2018: 128; Morgan, 2007: 51). Like this, Bryman (2006: 105) claims that mixed-method research employs both qualitative and quantitative research methods to address both "what is it like" and "how many" questions

(Johnson & Onwuegbuzie 2004; Creswell 2018). While qualitative assessments produce assertive data that provides precise details and aids in assessing the study's investigation goals, quantitative analyses use descriptive and inferential statistics. According to Ng'eno (2018: 88), qualitative and quantitative elements may be carried out sequentially or concurrently, emphasizing either component or equal weight being given to both.

4.4 Case Study

The Durban University of Technology (DUT) was created in April 2002 through the merger of Technikons ML Sultan and Technikon Natal. It was formerly known as the Durban Institute of Technology. Later, to maintain consistency with other UoTs' names and statuses, it changed its name to the Durban University of Technology. The six faculties of DUT, which include Accounting and Informatics, Applied Sciences, Management Sciences, Engineering, and the Built Environment, Health Sciences, and Arts and Design, are home to about 33 000 students. DUT is a technology university with multiple campuses and belongs to the International Association of Universities. It leads the way in advanced study, technological education, study, and innovation. As research at UoTs is advancing quickly, and some of its institutions are already listed in international university rankings that assess the quality and impact of academic output. DUT was listed among the top 500 universities in the world in 2020, coming in at number 10 for global citations and number 5 for national rankings (Durban University of Technology website 2023). In the categories of publications output units by publication type; relative cumulative share to sector research publications outputs by individual universities; percentage of book publication output units by a university; growth in weighted per capita research production; and normalized weighted research outputs, DUT is ranked as the top UoT in the Republic of South Africa, Department of Higher Education and Training Research Output Report (2022: 18). This study is vital because DUT is a thriving UoT in a field of research that was previously not a focus. It is crucial to comprehend the ecosystem of research data management that underpins ongoing research activities. The Faculty of Accounting and Informatics (FAI) was selected as a case study for the novel design model produced by this research study. The faculty selection was based on the significant presence of masters and doctoral candidates, including the researcher leading the study, primarily from the Library and Information Studies department. Understanding RDM techniques and existing research data management standards of the faculty was necessary. A research data repository platform model was designed using the study results, which evaluated the depth and breadth of RDM knowledge and practices.

4.5 Research Methods

The phases of the scientific procedures, tactics, and strategies utilized in research to find new knowledge or gain a deeper understanding of a topic are referred to as research methods. Quantitative, qualitative, and mixed methods research are the three most popular methods.

4.5.1 Quantitative Method

This method of inquiry is useful for figuring out how frequently or how much something exists. Creswell (2018: 153) claims that rigorous data quantification and tightly controlled scientific variables are the main components of quantitative research methodologies. The quantitative research approach makes identifying patterns or connections easier and creates generalizations. The quantitative research method sustains an empirical paradigm in which the research is independent of the researcher (Creswell 2018: 153). To measure reality objectively, data is used, and by demonstrating objectivity, quantitative analysis lends meaning to the data it gathers. This approach was used in the study because of the aforementioned crucial characteristics.

4.5.1.1 Survey Questionnaire

A web survey questionnaire was the primary tool for gathering information about postgraduate students' RDM practices. According to Babbie (2016: 270) and Connaway and Powell (2004: 146), a questionnaire contains a list of questions the respondents must fill out independently. In "any research involving human beings," questionnaires are frequently used to gather data (Pickard 2007: 183). Questionnaires are a standard and efficient tool used in many library and information studies in South Africa that examine user needs and evaluate services (Majyambere 2014; Muchaonyerwa 2015; Sejane 2017; Mutsvunguma 2019). Based on how the data is gathered, there are two types of questionnaires: self-completed and completed by an interviewer (Saunders, Lewis & Thornhill 2012: 373). The respondents fill out self-administered surveys divided into a browser or web-mediated, intranet-mediated, posting- or email-based, and delivery- and gathering-mediated questionnaires, depending on the distribution method (Saunders *et al*. 2012: 373). Contrarily, the researcher completes the forms that the interviewer fills out after reading the questions to the respondents and recording their responses. Depending on how the interviewer and respondent communicate, these surveys can be categorized as either structured interviews or telephonic questionnaires (Saunders *et al.* 2012: 373).

Online surveys served as a vital research tool during the COVID-19 pandemic, offering a quick and efficient way to collect data. In many contemporary research fields, gathering survey data on uniform survey questions sent to the target group or sample is a crucial data collection tool (Hlatshwako 2021: 76). Researchers should consider two different kinds of questionnaire questions; unrestricted questions, which give respondents the freedom to choose the phrasing and stretch of their responses and express their complex opinions; and closed-ended questions, for which the researcher has already predetermined the answers and the respondents are prompted to choose from the available options (Denscombe 2008: 271). The study objectives, the adopted models (DAF, CCMF, and UCRDMF), and the literature (see Annexure H) determined the questionnaire's substance. The survey comprised two portions (see Annexure H). Demographic and background data were presented in Section A. It has been discovered that demographic factors substantially impact how people use technology (Venkatesh *et al*. 2003: 456).

Research data classification, management; maintenance; and technological issues were covered in Section B. The sub-sections of section B were further separated according to the research objectives. Items like Identification, Classification, and Location of Research Data Practices addressed ascertaining research data management. Items like value and maintenance practices and research data reuse practices discussed the objective of establishing the usefulness of the faculty research data repositories. Items like designing research data repository platform prototype practices addressed the objective of designing an innovative digital prototype for a faculty research data repository platform based on research findings. Closed-ended questions may have been included in the questionnaires because the researcher presented the options, and the respondents chose the best ones that applied to them. The survey contained two categories of closed-ended questions, including yes-or-no inquiries, while other questions were presented as Likert scales or other numerical measurements. Since all respondents provided the same questions and responses, the researcher could evaluate the data consistently (Saunders *et al*. 2012: 375). The web questionnaire, administered using the QuestionPro software program, was chosen because of its potential to quickly gather enormous amounts of data from widely scattered populations. When respondents could not select one of the possible answers, they were asked to provide their responses by choosing the "other" option. The quantitative data made it easier to analyze with SPSS and allowed a rapid and straightforward interpretation of the findings. The master's and doctoral students who provided the data for this study had some background knowledge of RDM due to their research studies. The researcher realigned the quantitative data received through a questionnaire with the qualitative data obtained through interviews to avoid any biases resulting from the shortcomings of a questionnaire.

4.5.1.2 Reliability Testing of the Questionnaire

This study's questionnaire material was meticulously refined to guarantee the quantitative data's validity and reliability. The study used Cronbach's alpha for each of the survey questions to determine the level of consistency, adhering to Golafshani's (2003: 598) guidelines indicating that the degree to which a measurement is consistent when given repeatedly; the continuity of a measurement consistency over time; and the correlation of measures within a given period are the three kinds of validity used in quantitative research. Academics claim that Cronbach's alpha is widely accepted and that it provides a degree to which internal reliability of a magnitude or test; is represented by a value between 0 and 1" (Dennick & Tavakol 2011: 53). Results from tests between 0.7 and 1.0 are regarded as acceptable because they demonstrate the reliability of the data. The researcher calculated each questionnaire question's Cronbach's alpha, and any that fell short of the required 0.7 were improved, tested, and validated until the required 0.7 was reached.

4.5.1.3 Population and Sampling

As per Sekaran and Bougie (2010), the entire group of individuals, activities, or research subjects the investigator is interested in constitutes a study population. The candidates were drawn from the DUT Faculty of Accounting and Informatics. The study comprised 98 Masters and Doctoral-level postgraduate students registered between 2015 and 2020. The institution's faculty office provided the list. The postgraduate students were identified for their active roles in contributing to research data.

Sampling is crucial in research since it influences the results' reliability. Numerous scholars claim that random sampling has historically been linked to quantitative research, while nonrandom sampling has always been linked to qualitative research (Creswell and Plano Clark 2018). External validity refers to the ability of quantitative research sampling to choose analytical units characteristic of the population to permit the generalization of the findings. Probability sampling, which gives every member of the population an equal and independent chance of becoming part of the sample, can be used to accomplish this (Gelo, Braakmann & Benetka 2008: 274). Simple random sampling, structured random sampling, stratified random sampling, and cluster sampling are some of the most used methods for sampling probabilities.

The first stage of the study's population comprised 98 Master's and Doctoral-level postgraduate students registered between 2015 and 2020 who took the web survey. According to Leedy and Ormrod (2010: 213), the entire population should be surveyed for smaller groups (N = 100 or fewer). A census of 95 master's and doctoral-level postgraduate students who volunteered to participate was conducted because the study's population was so modest. Israel (2013) defines a census as a method that considers every component of the population. Three of the 98 postgraduate master's and doctoral students did not consent to the study or explain. The justification for using a census in this study is based on the recommendation that study sites under 100 should survey the entire population (Leedy & Ormrod 2010).

4.5.1.4 Quantitative Pilot Study

A pilot study involves a small group of specialists or intended audience members reviewing the data-gathering tool (Saunders *et al.* 2012: 420). To spot potential issues and address them before the data collection began, the researcher identified candidates who met the criteria for the web survey; five respondents met the criteria as Masters and Ph.D. students in the Faculty of Accounting and Informatics at DUT between 2015 and 2020. To confirm the candidates' involvement in the pilot project, a Consent Form and a Letter of Information were provided to each. Once permission was secured, a link to the pilot survey questionnaire was emailed to each participant. The email informed recipients that participation was optional, that their right to personal autonomy was respected, that they were assured confidentiality, and that no names would be listed on the survey. The online survey by QuestionPro served as the measurement tool for the pilot project.

The respondents concurred that the provided web questionnaire's questions and instructions were unambiguous, straightforward, and concise. In the respondents' opinion, all questions were pertinent, and no duplicates were found. Additionally, many believed the inquiries to be neutral and unnerving. No difficulties have been documented in the language and grammar used in the tools.

4.5.1.5 Quantitative Data Collection

The study used an online questionnaire to collect data from master's and doctoral postgraduate students. Before beginning the data-gathering process, the researcher had to contend with ethical clearance considerations, which necessitated obtaining and receiving authorization from university authorities to perform the study on campus. Using an online survey tool called QuestionPro, the researcher distributed the questions. The faculty office gave the researcher a list of master's and doctoral students enrolled between 2015 and 2020. In addition to the ethical clearance approval certificate and a link to the online survey, participants received an introduction letter asking for their permission.

4.5.1.6 Quantitative Data Analysis

In a research process, the analysis phase involves organizing, structuring, and making sense of a large amount of data obtained through various methods to answer the research questions. Bellamy (2012: 788) defines it as "the procedure of adjusting data, usually to find underlying patterns and answer research questions." The survey questionnaire's closed-ended questions were used to provide quantitative data. The researcher used percentages and mean and standard deviation tests to analyze the data. A QuestionPro survey was utilized to collect quantitative data, and SPSS version 28 was used to perform descriptive statistics in percentages and frequencies. The QuestionPro survey also exported graphs and charts.

4.5.2 Qualitative Method

Leedy and Ormrod (2010: 103) define qualitative research as a comprehensive approach that includes discovery. It gathers data about the individual's life events, feelings, actions, and significance. Sequential processes are standard in qualitative research as they occur naturally and enable the researcher to add depth by actively participating in the actual happenings (Creswell 2018: 115). As per Leedy and Ormrod (2010: 103), descriptive study is less organized in how it presents itself because it creates and constructs new hypotheses. Creswell (2018) asserts that inductive reasoning, instead of deductive reasoning, underpins qualitative research. The investigator's endeavor to justify is motivated by the observational elements that lead to queries.

A relationship is established between the viewer and the data in qualitative research as opposed to quantitative studies, which entirely isolates the investigator from the subject being investigated. (Creswell 2018). Data collecting involves gathering answers to questions from a pre-selected participant group. The research method entails data collection utilizing specific instruments to address the study's pre-established research topics (Dulle & Minish-Majanja 2011: 35).

4.5.2.1 Qualitative Data Collection

Interviews were employed in the study as a qualitative data-gathering tool, and master's and doctoral student supervisors at the faculty of Accounting and Informatics at DUT participated in interviews. Supervisors of doctorate and master's students were invited to participate in online interviews for the study. Zoom and Microsoft Teams video-conferencing platforms were utilized when conducting the interviews. Interviews were confirmed with supervisors based on their availability on suitable dates and times. Recording only commenced upon obtaining the supervisor's consent. Notetaking occurred throughout the process. The value of taking notes through interviews has been highlighted; Wahyuni (2012: 74) stresses that along with recording each interview, the researcher should also write notes during and right after each interview to capture supplemental data in the way of research transcripts. It has become simpler to record interviews on audio owing to the growth and adoption of ICTs, such as mobile technology applications. According to Denscombe (2008: 272), the audio clip of the interviews not only ensures the persistence of the recorded dialogue but also makes itself available for examination and validation by other researchers, reducing skepticism about the collected data. Before the interview, the researcher briefed the subject to describe the conversation's topic and emphasize the interview's voluntary, anonymous, and confidential character. During the subsequent debriefing, the researcher expressed gratitude to the participants for being part of the study and solicited their general feedback on the interview. During the debriefing, the researcher encouraged the participants to comment, ask questions, or add other details.

4.5.2.2 Interviews

The same concept has been preserved despite the use of varied vocabulary by numerous researchers to define research interviews. According to Babbie (2016: 271), interviews are a method of data gathering that can be conducted over the phone or in person between the interviewer and the study subject. Saunders *et al*. (2012: 372) define it as a deliberate dialogue between two or more people. It calls for the interviewer to build a connection with the subject, ask succinct, direct questions the subject is eager to answer, and pay close attention to what is being said. In an interview, the participant must interact with the interviewer and answer questions regarding a research problem. Using the interview approach, the researcher can access the components that would otherwise be inaccessible, such as independent individuals' different perceptions and opinions (Perakyla and Ruusuvuori 2011: 529). During an interview, the researcher and participant can freely converse, giving room for explanation and discussion. Another advantage of conducting interviews is comprehending the voices of participants' thoughts, viewpoints, and recollections. When necessary, the researcher might add new inquiries by confirming emerging themes and perspectives (Mertens 1998: 5). Saunders *et al*. (2012: 372) divide interviews into structured, semi-structured, and unstructured. In structured interviews, the interviewer presents the participants with questions and a list of potential replies (Pickard 2007). Data for this research were gathered using semi-structured interviews. Open-ended semi-structured interview questions enable gathering relevant data that may prompt a re-evaluation of the researched issues (Teddlie & Tashakkori 2009: 531). Since they combine elements of the structured and in-depth interview process, Saunders *et al*. (2012: 372) refer to these types of interviews as a hybrid in design. They use pre-established questions and themes to guide interviews but keep enough discretion to permit the respondent to respond extensively about any topic discussed throughout the interview (Wahyuni 2012: 74).

This study used interviews to obtain information from the master's and doctorate-level postgraduate students' supervisors in the FAI. The participants provided in-depth details on specific difficulties involving the RDM in their faculty. Through the interviews, the researcher elicited more justifications and instances from the participants, which thoroughly helped to comprehend supervisors' RDM concerns. The researcher developed and used an interview manual (see Annexure I). The interview guide considered these factors: open-ended main questions, follow-up questions, and probes (Wahyuni 2012: 74). To understand the specific themes, thoughts, ideas, and surprising viewpoints (Wahyun 2012: 74) regarding RDM methods, follow-up questions were used. Participants used prepared questions to keep the conversation going and clarify some debate issues by requesting additional information or supporting evidence for what has already been disclosed (Wahyuni 2012: 74). The interview guide should have an overview of the interview's objectives, a list of important questions, and a conclusion (Kvale & Brinkmann 2009). With the aid of the interview guide, the researcher maintained uniformity throughout the interviewees' responses (Boyce & Neale 2006). Using in-depth interviews, the researcher could gain a thorough understanding of each supervisor's feelings, ideas, and experiences regarding the RDM practices in the faculty (Pickard 2007). The researcher's material properly mirrored the participants' opinions, feelings, and perspectives. This supports Denscombe's (2008: 273) assertion that detailed interviews allow researchers to analyze the study's problem in-depth instead of only summarizing respondents' quick responses in a phrase or two.

These DUT FAI research participants' supervisors contributed substantial essential and inescapable knowledge for this investigation's success. Although interviews are recognized for their capacity to produce rich qualitative data in the social sciences, they have significant drawbacks (Wilkinson & Birmingham 2003). Wilkinson and Birmingham (2003) contend that collecting data through interviews with many participants is impractical. The interview sessions consisted of ten supervisors of master's and doctorate students in the FAI from DUT who participated in the study. Another challenge is the length of a single interview session (Wilkinson & Birmingha, 2003). The right amount of inquiries were made in the current study to ensure that one interview would not go longer than 50 minutes and that the questions would collect rich, pertinent, and complete data.

4.5.2.3 Document Analysis

It is feasible that some printed or digitally stored information gathered by another party for another reason could be used to fulfil the research needs of a specific study (Babbie 2016: 209). To evoke significance, gain a deeper understanding, and build empirical knowledge on the issue being studied, documents are reviewed and analyzed (Bowen 2009: 27). However, investigators are advised to approach records as merely indicative rather than as necessarily precise, accurate, or complete, and to determine their significance and relevance to the questions under consideration instead (Bowen 2009: 33). Including the records chosen for this research to support and complement data from the questionnaire and interview, results enable complete knowledge of the subject and the environment where it occurs (Yin 2014: 87).

A thorough analysis of pertinent studies on research repositories at global HEIs using the PRISMA method was carried out as part of the document analysis. The extensive search examined pertinent papers published in databases between January 2015 and March 2021 (Zibani *et al.* 2021: 241). The nine academic databases searched were Web of Science, SpringerLink Scopus, Google Scholar, SA e-Publications, Emerald Insight, Science Direct, and Ebscohost. The search produced a total of 2,000 records retrieved from the database. The remaining 75 actual article records were examined for significance using the eligibility requirements after removing 1,925 duplicate entries and articles (Zibani *et al.* 2021: 241). After removing records found to be irrelevant (40), the 35 full-text articles were evaluated for eligibility. After removing the 24 full-text articles that did not entirely address the study's scope, a descriptive synthesis of the 11 pertinent articles was also completed.

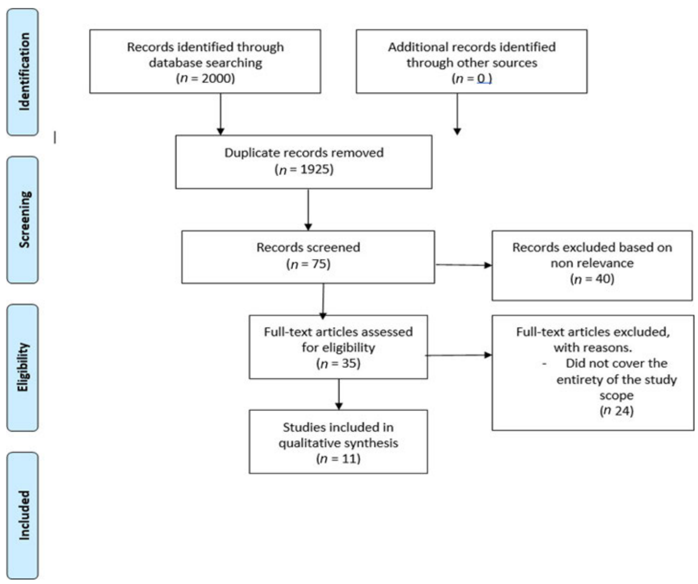

Figure 4.2. A flowchart for the PRISMA-guided systematic review

Source: Zibani *et al.* (2021)

4.5.2.4 Delphi methods

The Delphi method is a methodical and participatory research technique (Gollatta, Garza-Reye & Anosike 2018: 231-241). It is engaging because the researcher receives input from a panel of outside experts in a particular field. It is structured due to being compatible with quantitative, qualitative, and mixed method methodologies (Castillo-Megchun *et al.* 2021: 921). This allows the investigator to succeed in locating potential remedies for a conundrum. In addition, the Delphi method directs investigators in supporting research findings from investigations using mixed, quantitative, and qualitative techniques. As a result, it gives an identified research openness, significance, and path. The Delphi approach can be helpful when developing models and frameworks (Chowdhury, Katsikas & Gkioulos, 2022). Using the Delphi method, the researcher can obtain insightful feedback from professional experts in a relevant field and use that information to build a model or framework with reliable data. Furthermore, Delphi methods result in a sequence of phases for expert feedback (Gallotta *et al.* 2018). This implies that the Delphi method is applied recurrently until a research issue can be resolved by unanimous agreement.

A brief questionnaire can be created during this procedure to be distributed to participants who have been identified as subject-matter experts. Throughout the procedure, participants' confidentiality is protected. Additionally, the responses are examined, and participants receive feedback. If the experts cannot agree on approaching the research problem, the Delphi method is used again. Figure 4.3 below provides a graphic representation of how the Delphi method was contextualized for the objectives of this investigation. It demonstrates the methodical stages the researcher would take to work with specialists to determine how the faculty research data repository platform model suitable for a UoT should be created.

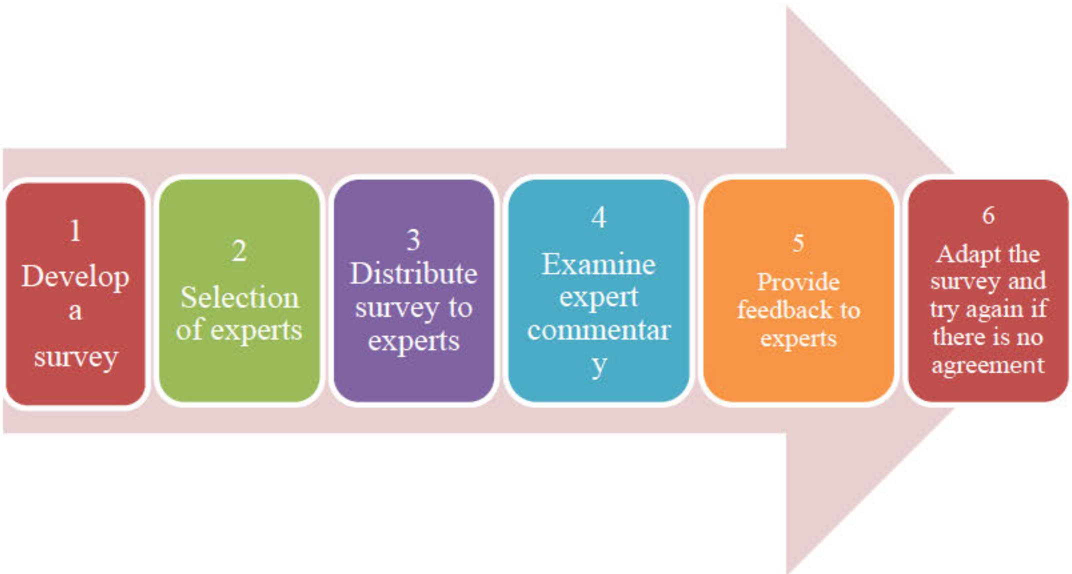

Figure 4.3. DELPHI methods in research

Source: Gallotta *et al*. (2018)

Regarding this study, the researcher determined that the Delphi method would be beneficial when creating a faculty research data repository platform suitable for a UoT context. This technique opens up a chance for communication with knowledgeable, experienced experts in research data repositories. Their knowledge, along with the data from both research tools that had been analyzed and the results of a meta-analysis of research repositories in HEIs, enabled the researcher to take a close-up look at the design of a faculty research data repository platform. As a result, the Delphi method ensured a thorough and essential contemplation when constructing the design model for a faculty research data repository platform suitable for a UoT context.

4.5.3 Qualitative: Reliability and Validity

By conducting pilot research, it was proven that the data-gathering instruments were reliable and valid. According to Tashakkori and Teddlie (1998: 82), reliability is the degree to which the outcomes of an assessment precisely convey the actual scope or standard of a concept. It concerns a measurement method's accuracy, repeatability, and comparability. The suitability and correctness of the research methods employed to get the answers are established by validity (Kumar 2010: 177). The study will respond to the issues it wants to tackle if it is trying to measure what it is meant for. In this study, the interview guide was meticulously refined to guarantee the validity and reliability of the results.

4.5.4 Qualitative: Population and Sampling

The interview population was postgraduate supervisors from the FAI. They were selected because they were established players in the faculty's postgraduate study and research administration. Sampling is a crucial step in research since it affects the reliability of the researcher's findings. The primary problem in sampling is choosing a sample using a method that guarantees population depiction to the best of the ability, without any bias (Burns 2000: 83). It is difficult to make sample decisions in a mixed-methods study like the current one since sampling strategies must be made for both the qualitative and quantitative components of the investigation (Onwuegbuzie & Collins 2007: 281). The deliberate sample technique, often known as non-probability sampling, is frequently employed in qualitative approaches (Schofield 2006: 30). Non-probability sampling prevents the researcher from ensuring or predicting that each population segment will be included in the study. Convenience, quotas, and purposeful sampling are common non-probability strategies (Leedy & Ormrod 2010: 211). In qualitative research, the researcher will deliberately select people who have experienced the main event under examination (Creswell & Plano Clark 2018: 415). Components from a subpopulation are chosen by convenience sampling depending on accessibility and research interests (Gelo *et al.* 2008: 275). As a result, the researcher employed these sample strategies while still adhering to their standard research methodology. For instance, purposive sampling was used in qualitative research.

The FAI postgraduate supervisors took part in the semi-structured online interviews. They were important in the faculty's research and administrative assistance for master's and doctoral students. Purposive sampling was chosen for this sample. The faculty office provided a list of 32 postgraduate supervisors; eight of the 32 declined, indicating that they did not provide supervision for masters and doctoral degrees. Five other people also declined, indicating no familiarity with research data management. Only ten of the nineteen remaining agreed and participated in the interviews. Babbie (2016: 94) defined purposive sampling as comprising the observational units based on the researcher's estimation of the most useful ones. Purposeful sampling was used in this study to obtain information from individuals directly influenced by RDM concerns and who, as a result, had access to rich material (Onwuegbuzie & Leech 2007).

4.5.5 Qualitative Pilot Study

To pilot the research with supervisors in the faculty, consent was sought for interviews. The information letter was forwarded to two supervisors at the DUT Faculty of Accounting and Informatics for the piloting objective. The participants were asked to comment on the questions and instructions that needed more clarification, identify irrelevant questions omitted or repeated; note the amount of time it took them to complete the questionnaire or respond to interview questions; identify questions that were unsettling or offensive; find ambiguous questions, and review the language and grammar used. In line with Creswell (2018), the researcher believes an instrument should obtain feedback from multiple participants, such as peers, scholars, researchers, and commentary from seminars and conferences.

4.5.6 Qualitative Data Analysis

Analysis of qualitative data was viewed by Gray *et al.* (2016: 269) as expressive, intricate, time-consuming, and thus costly. This is due to the excessive amount of data gathered using qualitative techniques. In this study, the data obtained through interviews were analyzed qualitatively. With the aid of the computer program NVIVO Pro, qualitative data were analyzed using content analysis through the thematic analysis approach. Codes were applied to generate themes. The study used a transparent, repeatable coding system that produced objective results (see chapter six, section 6.5). The first step in coding was to select a small number of precise labels (codes) for the transcript responses. As a result, coding was simple and could be finished quickly and easily. Data were coded according to the research questions, categorized by topic, and presented verbally and narratively. According to Weber (1990:15), "topic content analysis is the procedure of grouping many words in a text into smaller categories/themes." Four major themes emerged from the qualitative analysis, each with numerous subthemes. The following themes emerged from this study: theme 1: Importance of research in professional life; theme 2: Data attributes; theme 3: Data management; theme 4: Repository development. The study's findings are presented in chapter 6 based on the themes that surfaced during coding. The NVIVO software allowed the researcher to group and sort data and see relationships.

4.6 Ethical Considerations

Research ethics is the appropriateness of the researcher's behavior toward the study participants or other people who may be affected (Gray 2009). Informed permission, consequentialism (not harm), respect for incognito, secrecy, and privacy are a few of the most significant ethical considerations in research (Fouka & Mantzorou 2011: 7). The researcher essentially checks that the proper procedures have been followed to protect voluntary participation during this process. The participant's informed consent, institutional and ethical approval, and ongoing assurances of confidentiality and anonymity are all requirements. Ethical requirements were met for this research. The researcher successfully finished the curriculum in Introduction to Ethics (refer to Annexure G). Faculty Research Ethics Committee additionally provided ethical permission (see Annexure E), and the Institutional Research and Innovation Committee (IRIC) issued Gatekeeper Permission (refer to Annexure F).

4.6.1 Recruitment of Participants

An inquiry was made to the faculty office for Accounting & Informatics for a list of master's and doctoral students enrolled from 2015 to 2021. The participants received correspondence requesting their authorization to participate in the study. The pupils received the information letter and approval form via email. After receiving approval, participants received an email with a link to the survey. They were reminded that participation was entirely up to them. A list of supervisors was also obtained from the faculty. The same procedure was also used; supervisors were sent a letter with information and a request for consent. Following the approval of authorization, each supervisor received an invitation for an online interview via Microsoft Teams or ZOOM.

4.6.2 Informed Consent

Every participant received a letter with information about the study and a consent form (refer to Annexure C). Once permission was secured, participants were emailed a link to the QuestionPro-generated survey questionnaire.

4.6.3 Anonymity and Confidentiality

The researcher made sure the confidentiality of the participants was upheld. The survey did not include any name entries. The researcher reassured participants that all responses given during interviews would remain anonymous and secure and that all recordings and transcripts would be safely archived and then deleted after five years.

4.6.4 Protection of participants

The study's participants could withdraw at any time, and their participation was optional. The researcher ensured that participants' rights and welfare were protected and that they weren't subjected to abuse or victimization due to their involvement.

4.7 Appraisal of the Chapter

Chapter four provided a comprehensive overview of the research design and methodology employed in the study. The chapter began by describing the research design grounded in a pragmatic philosophical framework. Pragmatism was chosen as the philosophical foundation to guide the study's approach and methodology. The mixed methods research method was selected as the investigative approach, indicating a combination of qualitative and quantitative research strategies were employed to gather and analyze data.

Furthermore, the chapter outlined the development of data collection tools, drawing upon guiding principles from existing literature and theoretical frameworks. This emphasized the importance of a theoretical foundation and existing knowledge in shaping the research instruments. The questionnaires' reliability was evaluated using Cronbach's alpha score, demonstrating a measure of the internal consistency and reliability of the data collection tools.

The chapter also discussed the pilot study, which involved five responses from the intended audience—masters and doctoral students from 2015-2020. Additionally, qualitative data was collected through interviews with two supervisors, recognized for their significant roles in faculty mentoring and research administration.

In terms of data gathering, a combination of interviews with faculty research supervisors and a web survey distributed to masters and doctoral students was employed. This approach ensured a comprehensive understanding of the research context by gathering student and supervisor insights.

The chapter provided essential information on the population, sampling methods, data analysis techniques, and ethical considerations. This demonstrated a thorough and systematic approach to conducting the research, ensuring the collection of relevant and reliable data while upholding ethical standards.

Ultimately, Chapter four served as a foundational framework for the subsequent chapter, where the findings derived from this robust research design and methodology will be reported and analyzed.

CHAPTER FIVE: PRESENTATION OF QUANTITATIVE RESULTS AND DISCUSSION

5.1 Introduction

In this chapter, the survey questionnaire's findings are presented in conjunction with a review of the results. The questionnaire served as the main instrument for data collection. The questionnaire was distributed to postgraduate students, namely masters and doctoral students in the FAI from the DUT enrolled from 2015 to 2020. Creswell (2018: 109) describes data presentation and evaluation as steps researchers take to summarize the main ideas from their research, followed by a discourse on the results. The results are displayed in graphs, cross-tabulations, and other figures for the quantitative data gathered. While results from correlations and chi-square tests where inferential methods and p-values were used to interpret them. The conventional method of documenting an outcome necessitates a statistical relevance explanation. Using a test statistic, a p-value is produced. With "p 0.05," a result is considered significant (Grabowski 2016: 4). The researcher used the explanatory sequential mixed method study diagram shown in Figure 5.1 to provide a graphic overview of the chapter contents, points of discussions, and findings.

Explanatory Sequential Mixed Method Study

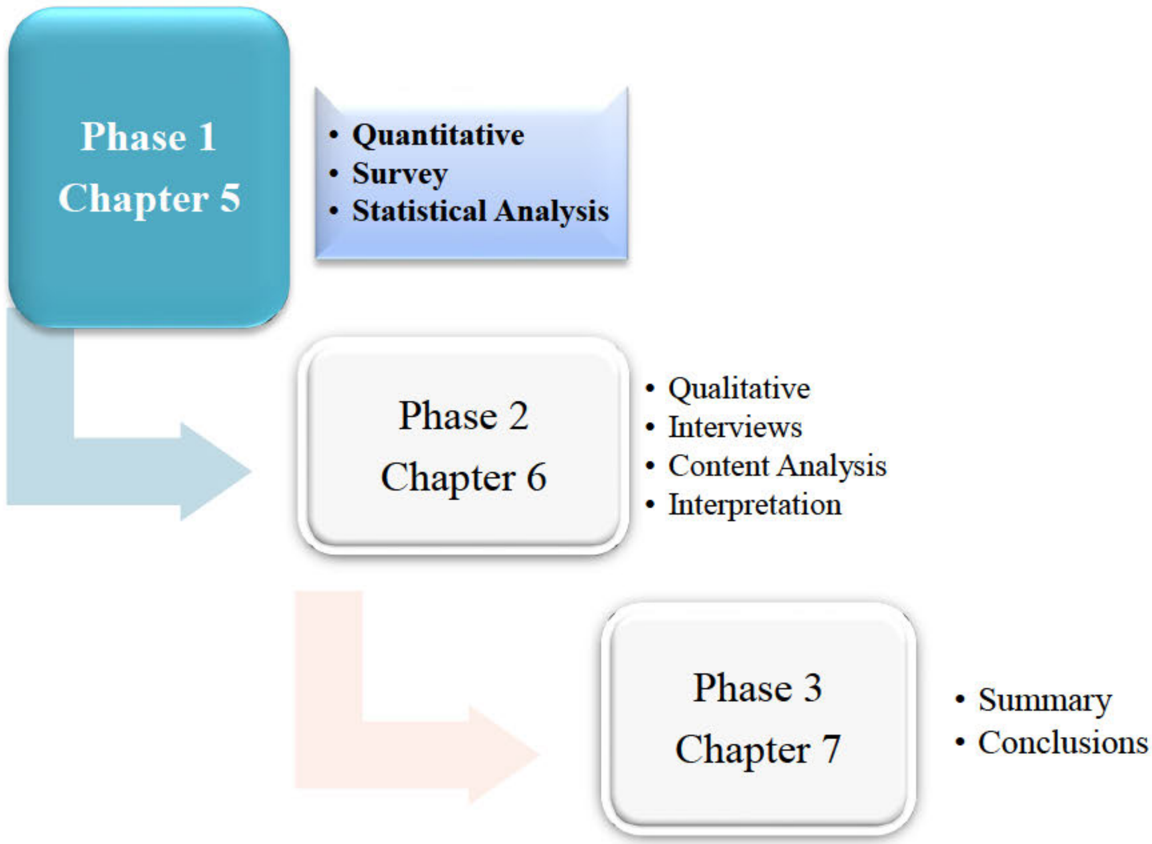

Figure 5.1. An adaptation of Creswell Explanatory Sequential Mixed Method Study

Source: Creswell and Plano Clark (2018)

5.2 Analysis of quantitative data from the sample

The study's purpose was to investigate how faculty members at the University of Technology manage their research data. Data was gathered from the Faculty of Accounting and Informatics at the University of Technology under study using a web questionnaire relating to biographical, data classification, management, maintenance, and technical aspects. In total, 95 questionnaires were dispatched, and 68 were returned, giving a 71.5% response rate. For such surveys, a 100% response rate is frequently difficult. According to Babbie (2016: 261), academic research should typically have a response rate above 55.6%, though follow-ups and the sort of questionnaire delivery technology used can both boost response rate success. A response rate of 60% is considered reasonable in the social sciences. The researcher had to remind the participants three times to get the 71.5% response rate.

5.2.1 The Research Instrument

There were 93 items in total in the research instrument, each with a nominal or ordinal level of measurement. The survey consisted of 24 questions that measured different themes.

5.2.2 Reliability Statistics

As indicated in the preceding chapter, the study's questionnaire instrument was meticulously refined to guarantee the validity and reliability of the results. Validity and reliability are the two most crucial components of accuracy. Multiple assessments on the same areas of study are taken to calculate reliability. While 0.70 is generally accepted as an acceptable value, Hair *et al*. (2016) state that values as low as 0.60 might be appropriate for exploratory research. Similarly, Saunders *et al.* (2012) concur that a newly developed concept is deemed "acceptable" when the reliability coefficient is 0.60 or higher. George and Mallery (2003: 231) also recommend a tiered system consisting of the following categories: ".9 - Excellent,.8 - Good,.7 - Acceptable,.6 - Questionable,.5 - Poor, and 5 - Unacceptable." Table 5.1 reflects Cronbach's alpha score for all the questionnaire items.

Table 5.1. Reliability Statistics

	Section	Number of Items	Cronbach's Alpha
Q15	Factors that deter students from sharing research data	13	0.817
Q16	Conditions for research data sharing	6	0.715
Q18	Factors that discouraged research data reuse	9	0.795
Q23	A Design for a faculty research data repository platform based on research evidence	4	0.868

Each section's reliability scores are higher than the alpha value suggested by Cronbach's alpha. This demonstrates that the grades for these research sections were generally accurate and suitable.

5.3 Biographical Data

Lyon *et al.* (2012) underscore the significance of comprehending RDM stakeholder characteristics in the CCMF because they are crucial for assessing how prepared communities are to carry out RDM activities. The researcher examined participant demographic information, including gender, departmental affiliation, and educational background. This section summarizes the respondents' characteristics, such as departmental affiliation, gender, and qualification, as requested in Section A of the questionnaire (See Annexure H). Table 5.2 describes the gender distribution of postgraduate students in the UoT faculty under study.

Table 5.2. Gender distribution

	Frequency	Percent
Male	28	41.2
Female	40	58.8
Total	68	100.0

The ratio of males to females is approximately 2:3 (41.2%: 58.8%) ($p = 0.146$). Female respondents comprised 58.8% of the total respondents, compared to male respondents, who included 41.2%. This is consistent with a Statista report from 2020, which confirmed that approximately 29, 7 million women and 28, 86 million men were living in South Africa in 2019. The study by Chawinga (2019), which aimed to comprehend research data management in Malawian universities, contrasts this finding. The majority of his research participants were men who predominated in higher education in Malawi. Table 5.3 indicates the academic disciplines to which the respondents belonged.

Table 5.3.Academic disciplines demographic data of respondents

	Frequency	Percent
Library and Information Studies	24	36.9
Financial Accounting	10	15.4
Information Technology	7	10.8
Information Systems	4	6.2
Office Management	4	6.2
Corporate Management	4	6.2
Information and Corporate Management	3	4.6
Auditing	1	1.5
Taxation	1	1.5
Financial Systems	2	3.1
Business Systems	1	1.5
Department of Information Systems	3	4.6
Financial Management	1	1.5
Informatics	1	1.5
Information Management	1	1.5
Office Technology	1	1.5
Total	**68**	**100.0**

The Library and Information Science, Accounting and Information Technology respondents constituted approximately two-thirds of the sample (63.1%). Library and Information, Science and Information Technology have masters and doctorate programmes. Hence this accounts for the high postgraduate student numbers in these programmes. Table 5.4 indicates the highest qualification that respondents were pursuing.

Table 5.4. Respondents' qualifications

	Frequency	Percent
Masters	58	85.3
PhD	10	14.7
Total	**68**	**100.0**

Most respondents pursued a master's degree, followed by those with a Ph.D. The findings were further supported by a statistical p-value test, which demonstrated that there were variances in credentials pursued that were statistically significant. There were significantly more respondents pursuing a master’s qualification (85.3%) than with PhDs ($p < 0.001$).

5.4 Ascertaining the management of the existing research data practices

Higgins (2011) asserted that knowing the data's origin is crucial for correctly understanding research data management processes. The DAF (Jones *et al*. 2009) and CCMF Lyon *et al*. (2012) integrate data management activities to emphasize the data life cycle and its significance in the research. To determine the identification, classification, and location of existing research data, Q4–Q11 in the questionnaire were used (see Annexure H). These questions revealed characteristics and circumstances deemed advantageous for ensuring proper research data management; the results are presented in the sections below.

5.4.1 Data attributes

This theme examined the attributes of the existing research data produced by postgraduate respondents related to research management. The primary research data types depended on the type of research study conducted. Figure 5.2 indicates the four types of research data generated by the faculty under study.

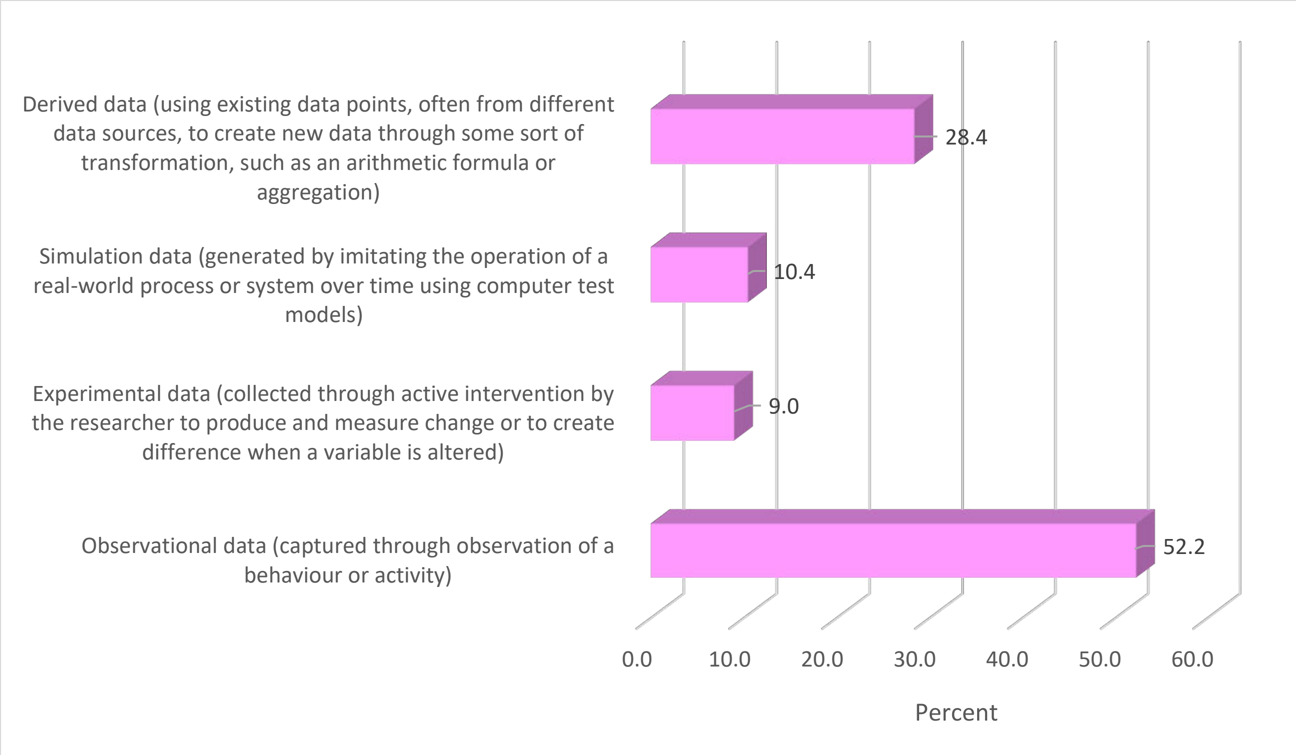

Figure 5.2. Research data produced by postgraduate students

The results revealed the following research data types: observational data, 35 (52.2%); derived data, 19 (28.4%); simulation data, 7 (10.4%); and experimental data, 7 (9.0%). The conceptual part of the Data Audit Framework (DAF) (2009) places emphasis on the value of auditing data assets to effectively manage data holdings. Assessing data attributes informs the faculty under study of the nature of the data they possess, the state, and its worth.

5.4.2 Research data used and produced by students.

The most significant amount of research data the respondents produced was mentioned in this item. Figure 5.3 indicates the amounts of data used and produced by students.

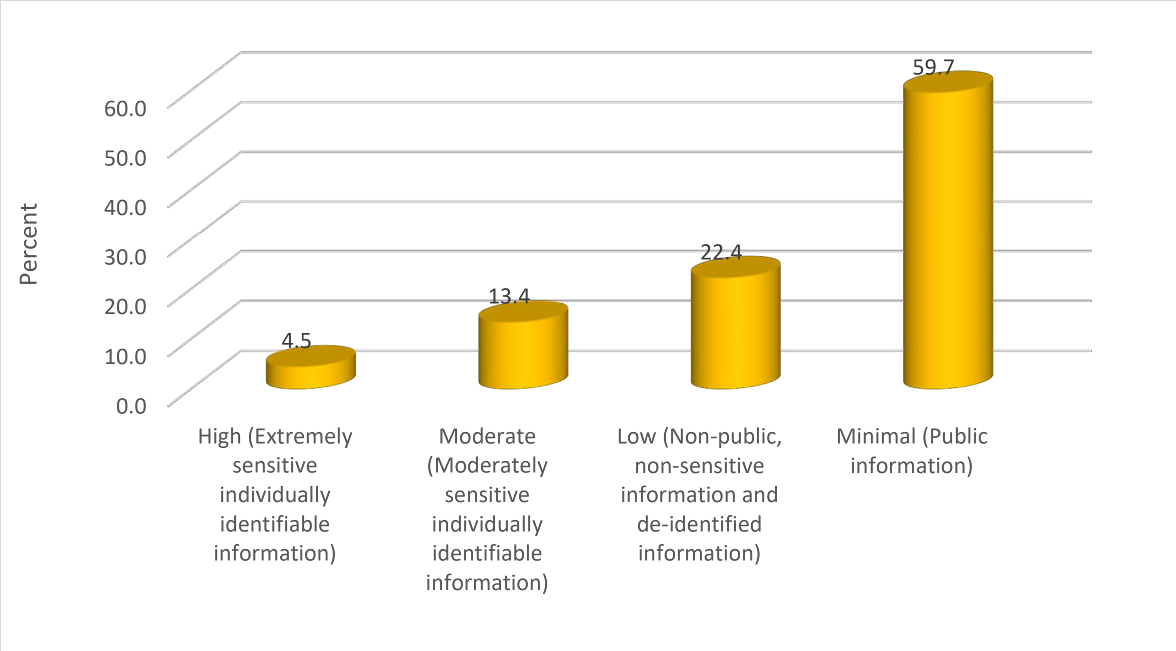

Figure 5.3. Research data used and produced by students

Most respondents used minimal (public information) data (59.7%), with a small number producing high (extremely sensitive individually identifiable information) (4.5%) ($p < 0.001$).

The data showed that most of the research data created by students in the faculty under study are public information (59.7%, N=68), requiring little to no classification. The next lowest category was non-public, non-sensitive, requiring some de-identification (22.4%, N=68). Just above (13.4%) were the moderate and high (4.5%), respectively moderately sensitive, exceedingly sensitive, and immediately recognizable. This analysis concludes that most student-generated material belongs to the least sensitive classification group.

5.4.3 Research data storage and backup

Respondents were questioned about their choice of digital storage services for storing the data from their studies safely and backed up. Figure 5.4 indicates the respondents' devices for backing up and storing data. There was room for multiple answers.

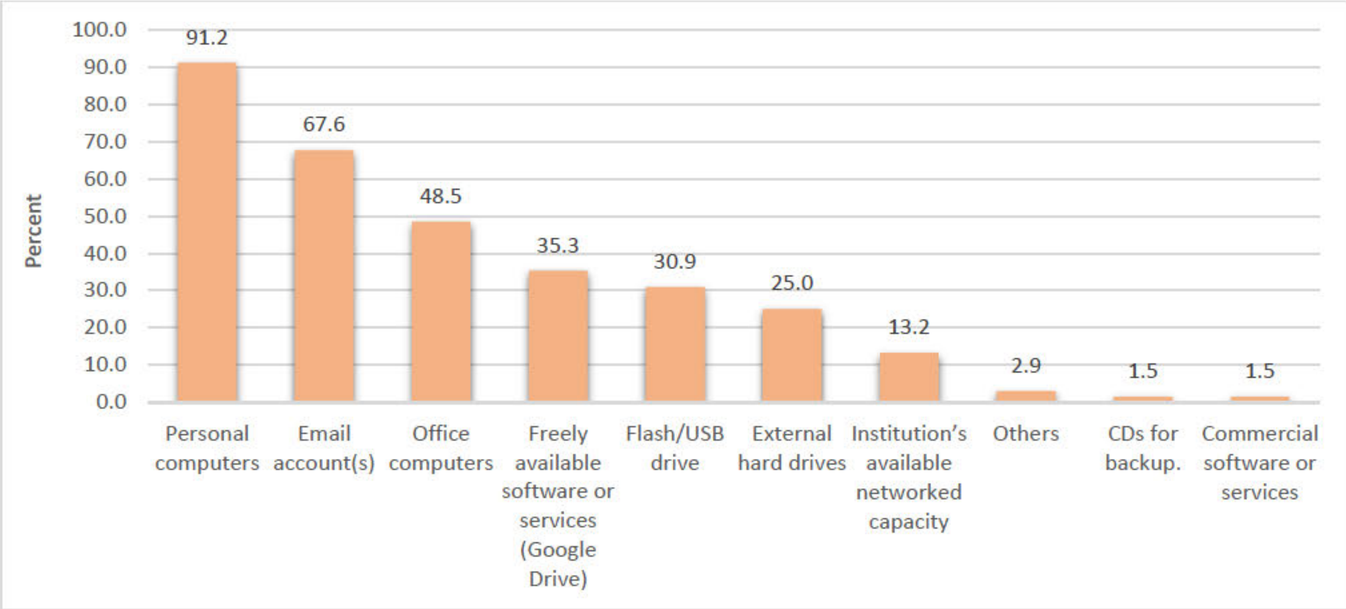

Figure 5.4. Data storage choices of students

According to the current study, researchers tended to save their data on personal, email accounts, and office computers (91.2%, 67.6%, and 48.5%, respectively). Students were more likely to use free software or services (35.3%), such as Google Drive. Findings showed that 30.9% of students stored study data on their devices or USBs. Only 13.2% of the students had data saved on the central servers of the university. In addition, it was discovered that another popular way for students to store digital files is on external hard drives or CDs (1.5%). The findings confirm those of previous studies by Bugaje (2019), Chawinga (2019) and Patterton (2016), which found that the majority of research data were stored on personal and business computers, via emails, external hard drives, and flash drives. Based on the analysis, the FAIR (Findability, Accessibility, Interoperability, and Reuse of digital assets) principles are not applied (Wilkinson *et al* 2016: 4). Research data remains hidden from the community that could benefit. The literature generally accepts that digital data are susceptible to loss or corruption due to software and hardware malfunction; therefore, regular backups are essential and can help identify hardware issues earlier (Consultative Committee for Space Data Systems 2012: 5; Cox and Pinfield 2014: 300).

5.5 Research data file formats, size, and management practices

This section provided students' perspectives on the formats, size, and management practice of pre-existing research data from their research projects.

5.5.1 Types of data employed by students

The participants were required to list the different types of data formats used to create research data for the study. The data types the participants used are depicted in Figure 5.5.

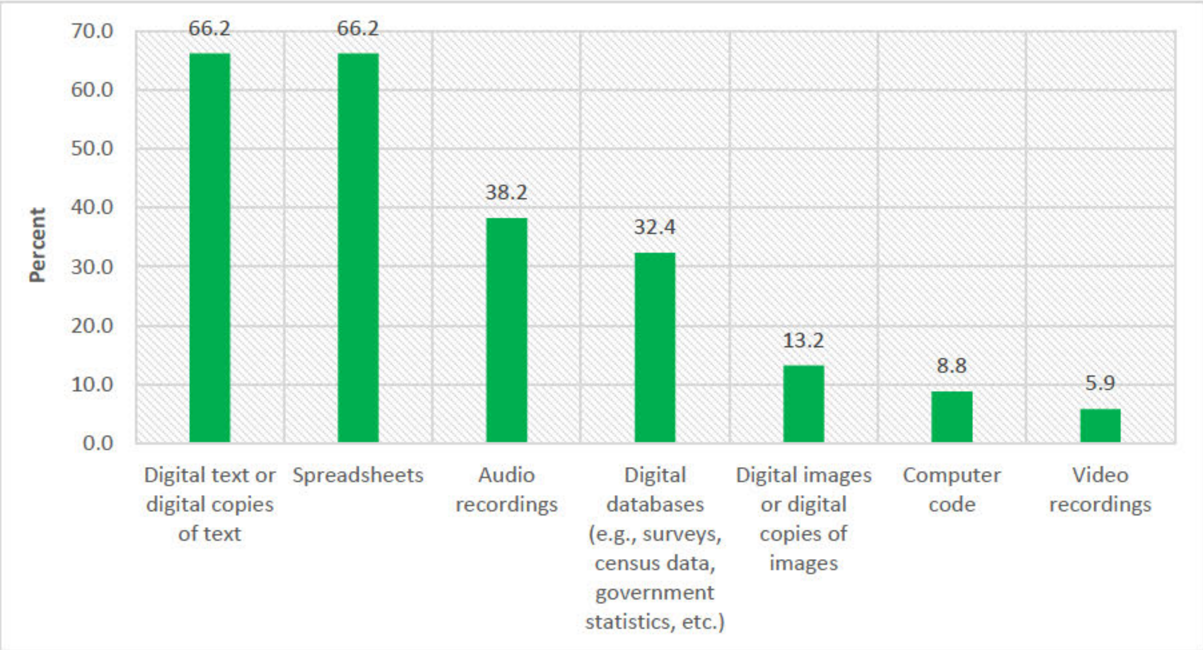

Figure 5.5. Types of data employed by students (N=68)

The system's capabilities for data discovery or presentation are determined by the attributes of the data, such as its size, type, or format. Understanding data properties is crucial for managing research data because data might differ significantly throughout study disciplines (Ohaji 2016: 25). The survey found that master's and doctorate students in the faculty under study frequently employed seven types of research data formats. The students' most commonly utilized and created file types appeared to be digitized or copies of text and spreadsheets (66.2% and 66.2%, respectively, n=68), shown graphically in Figure 5.5, followed by audio recordings and digital databases (38.2% and 32.4%), respectively. The faculty used fewer digital picture formats (13.2%), computer programs (8.8%), and video recordings (5.9%). Text documents are often small and do not take up a significant amount of space. Participants' digitally generated data supports the claims made by Chawinga (2019), Cox, and Pinfield (2014) that the rapid development of ICTs in academic and research settings has fueled the vast digital production of research data. The prevalence of audio recordings across human subjects' research domains and the prevalence of specimens in the medical sciences are unsurprising. However, just a small percentage of study data created required more than 100 GB.

Further analysis showed that participants produced less research data in artwork and video recordings. Studies by Chawinga (2019), Chen and Wu (2017), and others reported that none of their participants used video recordings as data forms. An ethical conundrum arises from video recordings used in research involving human subjects since many subjects might not be open to capturing their reactions on camera. As a result, video data in human research is not very common. The CCMF cautions that certain ethical requirements may restrict how researchers can use the data outside the primary objective for which agreement was sought and obtained from the participants (Lyon *et al*. 2012: 132).

5.5.2 Size of the research data that was produced

The amount of research data produced in the respondents' research projects was asked to be estimated. The size/volume of research data produced through the research project is shown in Table 5.5.

Table 5.5.Respondents' data size (N=68)

	Frequency	**Percent**
Less than 1 GB (gigabyte)	40	58.8
Larger than 1 GB but smaller than 100 GB	15	22.1
I don't know	10	14.7
More than 100 GB but less than 1 TB (terabyte)	3	4.4
Total	**68**	**100.0**

Approximately 59% of the respondents used less than 1 GB, with less than 5% using more than 100 GB ($p < 0.05$). According to the findings, 40 (58.8%) respondents claimed that their research projects generated 1 GB or less data. Additionally, 15 respondents (22.1%) mentioned that they produced data roughly over 1 GB but less than 100 GB. About 10 (14.7%) respondents were unsure of the volume of research data they produced for their projects. Additionally, 3

(4.4%) respondents produced in their research projects an average amount of data greater than 100 GB but less than 1 TB (terabyte).

5.5.3 Frequency of research data backed up

It is crucial to back up research data to reduce data loss due to theft, equipment damage, hard drive failure, or accidental deletion. Due to the high value, uniqueness, or difficulty of duplicating research data, backup strategies should be implemented as soon as possible. Protecting research datasets through storage and backup is crucial. The frequency of data backups is shown in Figure 5.6.

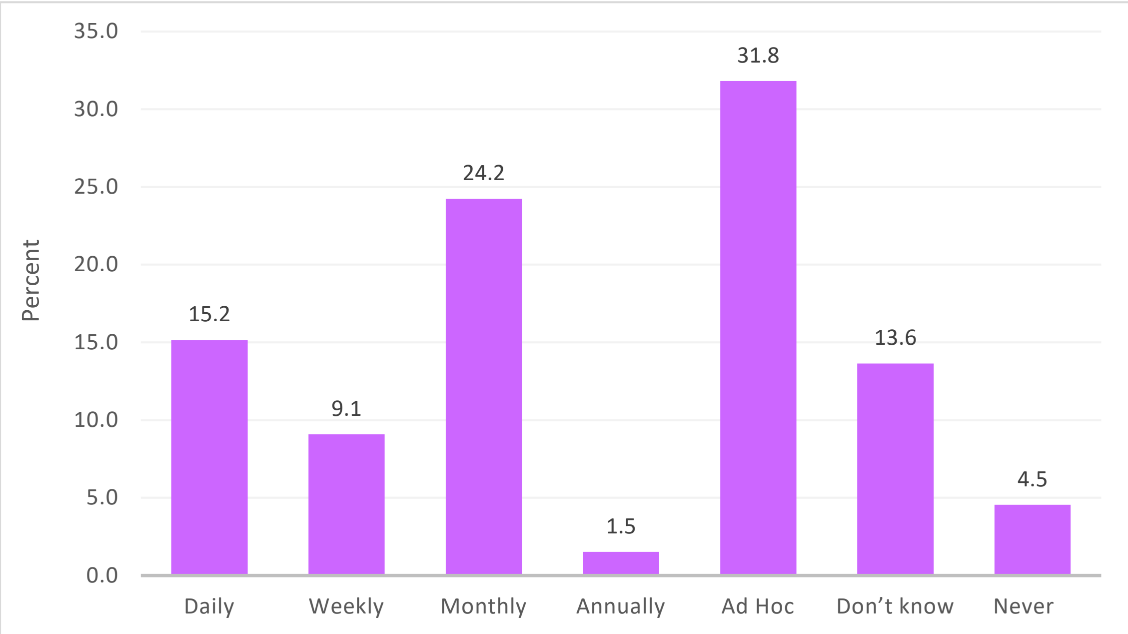

Figure 5.6. Research data backup frequency

There are significantly different data backup frequency levels ($p < 0.001$). According to the data, ad hoc (31.8%) forms of backup were the most typical response by the students. This is possible because backing up research data is still primarily a personal option and decision because there are no procedures or compliance standards. The data showed that daily backups were used 15.2% of the time, with monthly backups coming in second with 24.2%.

Further inquiry was made to understand respondents’ backup frequencies on their storage devices. The frequency of backup per storage device is demonstrated in Figure 5.7.

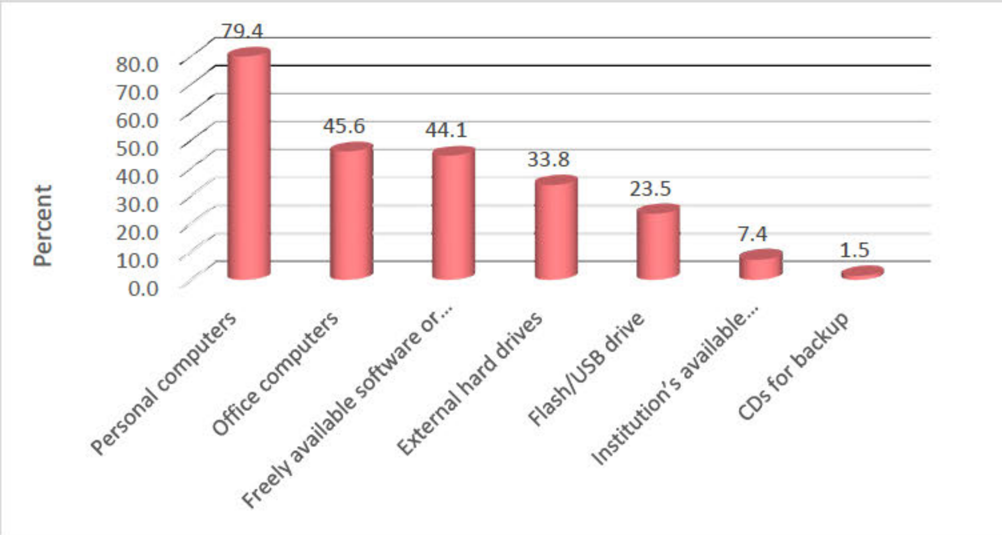

Figure 5.7. Backup frequency on devices

The results showed backup frequency on personal computers (79.4%), office computers (45.6%), freely available software or services (44.1%), external hard drives (33.8%), Flash/ USB drives (23.5%), institution's available networked capacity (7.4%) and CDs (1.5%). Multiple responses were permitted for this question. The data representation aligns with respondents' research data storage devices, where the backup is performed. As alluded that, it is a common occurrence in the literature and the study that digital data are vulnerable to corruption or loss as a result of hardware and software issues regular backups are crucial and can aid in the earlier detection of hardware issues (Consultative Committee for Space Data Systems 2012: 5; Cox and Pinfield 2014: 300).

5.6 Usefulness and maintenance practices of the faculty research data

Disciplines vary in their approaches to data sharing. If authors are prepared to give their data sets, then data can only be viewed, used, and reused. Molecular biology and ecology, for instance, have robust data-sharing cultures, but chemistry and history do not (Velden and Lagoze 2009: 675). These repositories offer more data that scientists are more likely to discover and use if they are firmly rooted components of the disciplinary research ecosystem (Mannheimer, Sterman & Border 2016: 8). In this context, the study explored data-sharing behaviours, emphasizing data-sharing motivational factors, technologies, and factors affecting data-sharing.

5.6.1 Research data sharing

According to Jeng and He (2022: 1281), data producers must be ready to disclose their data sets for the information to be accessed, used, and reused. Therefore, the study focused on the ecosystem of data-sharing practices, particularly the data-sharing instruments, motivators, and influences. The students were prompted to indicate whether they had shared the data they produced with other researchers and research stakeholders. Table 5.6 displays the results. The results demonstrate that 39 students (57.4%) responded with a yes and 29 (42.6%) with a no.

Table 5.6. Research data sharing frequency

	Frequency	Percent
Yes	39	57.4
No	29	42.6
Total	68	100.0

There is no discernible difference between the participants who shared and those who did not ($p = 0.0225$).

5.6.2 Motivating elements for students to share research findings

Disseminating papers or research findings is nothing new for scientists; however, it is now difficult for them to do the same with their data (Guedon 2015: 85; Bangani & Moyo 2019: 12). Students were asked to list the aspects that motivated them to share the research data they produced. The reasons for respondents disseminating information are shown below. Multiple responses were permitted for this question.

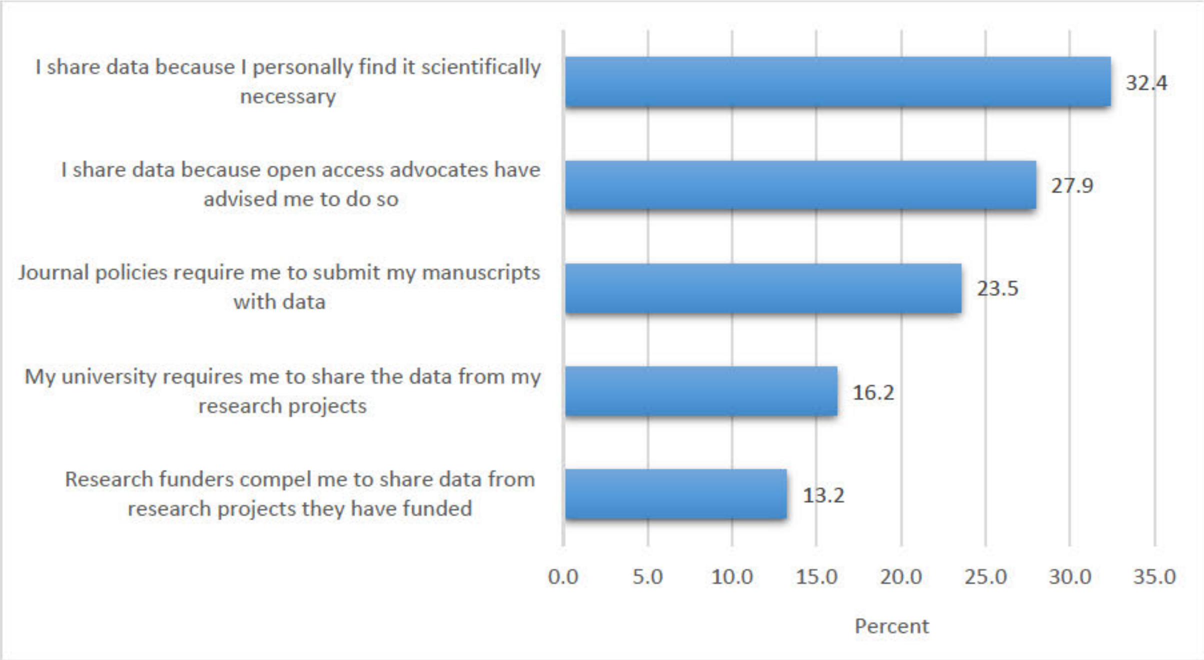

Figure 5.8. Research data-sharing elements (N=68)

Figure 5.8 shows that while 32.4% of students were motivated to share scientific research because it is necessary, 27.9% shared data because open-access advocates advised them to do so. Additionally, according to students, regulations of journals requiring the submission of submissions with data frequently influenced factors (23.5%). The least number of students (16.2%) indicated that university requirements were the direct impact that motivated submission. These findings imply that students' willingness to share data was primarily inspired by their activities, followed by open-access advocacy. For other scientists and health professionals to obtain the most recent evidence, use it to further their scholarship, and benefit from the expertise, open access movements promote open data access (Ng'eno 2018: 207). Guedon (2015: 85) believes that while scientists have traditionally shared their papers or research results, they must learn how to communicate their data. According to the Berlin Declaration (2003) and the European Commission (2012), research data storage in open-access data repositories is the only quick and dependable method to guarantee access to research data. These results confirm earlier studies' findings and, in some respects, go against them. For instance, even though research funders are praised for pressuring scientists to publish their data in studies Chen and Wu (2017: 346); Chawinga (2019), this factor had a minimal bearing on the current research.

5.6.3 Tools for sharing data

The following must be remembered: researchers' desire to exchange data with their counterparts or others is only possible with access to the appropriate data-sharing infrastructure. The complex technical infrastructure that Lyon *et al*. (2012: 133) included in the Community Capability Model Framework (CCMF) model includes several data-sharing mechanisms. The CCMF model states that technical infrastructure is required for collaboration, data exploration, access, and preservation. Researchers can share data in various ways, such as by linking data sets to publications that have already been published; keeping data sets in repositories; posting data on a website for individual or research facilities or responding to inquiries from other scholars for data (Wallis *et al*. 2013: 2).

A list of different data-sharing platforms was provided to respondents for this question, and they were asked to rate how much they had utilized them to disseminate information about the research data they had produced. A four-point Likert scale was used for the question, with the options being all, most, some, and none. The respondents who chose ALL indicated that they exchanged all their data using that specific data-sharing tool, and the respondents who selected MOST stated that they primarily used that specific data-sharing tool to share their data. The respondents who selected SOME indicated that they shared some data using that specific sharing tool. The respondents who selected NONE stated that they had never used that specific data-sharing tool. The scoring trends are outlined in Table 5.7.

Table 5.7. Scoring patterns for research data-sharing tools

	None		Some		Most		All		Chi-Square p-value
	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Through external drives (flask disks)	3	17.6%	10	58.8%	2	11.8%	2	11.8%	0.015
Through emails	2	6.3%	16	50.0%	7	21.9%	7	21.9%	0.005
Through journals websites	3	11.1%	16	59.3%	5	18.5%	3	11.1%	0.001
On social networks	7	53.8%	2	15.4%	4	30.8%	0	0.0%	0.232
On my website/blogs/wikis	8	72.7%	0	0.0%	2	18.2%	1	9.1%	0.020
Through clouds (Google Drive, DropBox, etc.)	3	16.7%	11	61.1%	2	11.1%	2	11.1%	0.005

University repositories	2	11.1%	1	5.6%	8	44.4%	7	38.9%	0.042
Through research funders' websites	5	33.3%	4	26.7%	3	20.0%	3	20.0%	0.865
On my faculty's website	7	58.3%	0	0.0%	3	25.0%	2	16.7%	0.174
On the principal investigator's website	4	44.4%	1	11.1%	1	11.1%	3	33.3%	0.392
Through a national network	4	44.4%	2	22.2%	1	11.1%	2	22.2%	0.550
Through a regional network	5	62.5%	1	12.5%	2	25.0%	0	0.0%	0.197
Through a global network	5	55.6%	2	22.2%	1	11.1%	1	11.1%	0.189
Other	5	100.0%	0	0.0%	0	0.0%	0	0.0%	-

Table 5.7 shows respondents' tools to share the research data they generated. Regarding sharing research data through external drives (flask disks), 17.6% respondents indicated none, while 58.8% indicated some, 11.8% respondents indicated most, and similarly, 11.8% indicated all. In terms of exchanging research data through emails, 6.3% responses were none, with 16 50.0% responses as some, while 21.9% responses were most, and similarly, 21.9% responses as all. Regarding sharing through e-journal websites, the results showed 11.1% of responses as none, 59.3% responses as some, 18.5% responses as most, and 11.1% as all. In terms of sharing data on social networks, 53.8% responses were none, 15.4% responses as some, 30.8% responses as most, and 0 responses for all. Regarding sharing data on personal websites/blogs/wikis, 72.7% responses were none, 0 responses as some, 18.2% responses as most, and 9.1% responses for all. Sharing data through clouds (Google Drive, DropBox, etc.) results showed 16.7% responses as none, 61.1% responses as some, 11.1% responses as most, and similarly, 11.1% responses for all. Regarding sharing data through university repositories, results showed 11.1% responses as none, 5.6% responses as some, 44.4% responses as most, and 38.9% responses as all. Sharing data through research funders' websites, the respondents' data showed 33.3% responses as none, 26.7% responses as some, 20.0% responses as most, and similarly 20.0% responses as all. With regards to sharing data on the faculty's website, the responses were in this manner; 58.3% were for none, 0 for some, 25.0% for most, and 16.7% for all. Concerning sharing data on the principal investigator's website, the responses were in this manner; 44.4% were for none, 11.1% were for some; similarly, 11.1% were for most, and 33.3% were for all. Sharing data through a national network presented responses in this

manner; 44.4% were for none, 22.2% were for some, 11.1% were for most, and 22.2% were for all. Sharing data through a regional network yielded responses in this manner; (62.5%) indicated none, 12.5% showed some, 25.0% was for most, and 0 was for all. Sharing data through a global network presented respondents' data in this manner; 55.6% were for none, 22.2% for some, 11.1% for most, and similarly 11.1% for all. The following patterns were observed:

- Some assertions exhibit higher levels of one option than others; in this case, the use of emails and journal websites to share data are such statements with responses as indicated above. Additionally, Rowhani-Farid and Barnett (2016: 5) noted that most scientists sent their data to one another via email, suggesting that emails are a standard data-sharing method. Sharing through journal websites has also become a standard compliance model associated with conditions of acceptance and publishing of an author's manuscript.
- In other instances, the options are similar; in this case, sharing data through the national network and on the principal investigator's websites represents such statements.

A chi-square goodness-of-fit test determined whether there was a significant difference in the scoring patterns for each statement and each option. The highlighted significant values (p-values: 0.015; 0.005; 0.001; 0.020; 0.005; 0.042) are less than 0.05, implying that the distributions differed. Therefore, the differences between respondents' scores (none, some, most, all) were significant.

Table 5.8. Scoring patterns for research data-sharing tools

	None		Some		Most		All		Chi-Square p-value
	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Through external drives (flask disks)	3	17.6%	10	58.8%	2	11.8%	2	11.8%	0.015
Through emails	2	6.3%	16	50.0%	7	21.9%	7	21.9%	0.005
Through ejournals' websites	3	11.1%	16	59.3%	5	18.5%	3	11.1%	0.001
On social networks	7	53.8%	2	15.4%	4	30.8%	0	0.0%	0.232
On my website/blogs/wikis	8	72.7%	0	0.0%	2	18.2%	1	9.1%	0.020

Through clouds (Google Drive, DropBox, etc.)	3	16.7%	11	61.1%	2	11.1%	2	11.1%	0.005
University repositories	2	11.1%	1	5.6%	8	44.4%	7	38.9%	0.042
Through research funders' website	5	33.3%	4	26.7%	3	20.0%	3	20.0%	0.865
On my faculty's website	7	58.3%	0	0.0%	3	25.0%	2	16.7%	0.174
On the principal investigator's website	4	44.4%	1	11.1%	1	11.1%	3	33.3%	0.392
Through a national network	4	44.4%	2	22.2%	1	11.1%	2	22.2%	0.550
Through a regional network	5	62.5%	1	12.5%	2	25.0%	0	0.0%	0.197
Through a global network	5	55.6%	2	22.2%	1	11.1%	1	11.1%	0.189
Other	5	100.0%	0	0.0%	0	0.0%	0	0.0%	-

Similar respondents allegedly gave similar scores for each statement's various options, according to the null hypothesis. The alternative asserts that there is a sizable distinction between the multiple options. Table 5.8 above illustrates the significance of the variations. The findings showed that there were statistically significant differences between the national network (chi-square of 2.111; dif (3) p-value = 0.550), a regional network (chi-square of 3.25; dif (2) p-value = 0.197), a global network (chi-square of (4.778) dif (3) p= value 0.189), and the website of the principal investigator (chi-square of (3) dif (3) p-value = 0.392).

The introduction of internet tools like emails, web portals, websites, data repositories, and collaboration platforms, among others, has transformed the sharing of data. The findings agree with the CCM framework, which elaborates on essential factors in data sharing, including formats for storing data, strategies for processing it, packaging, transfer policies, and procedures, data descriptions, vocabularies, semantics, and ontologies as data indicators (Lyon *et al.* 2012: 133).

5.6.4 Factors that deter students from sharing research data

This question assessed the degree to which certain circumstances made respondents less likely to share the study data they produced with other researchers. Respondents had to indicate on a Likert scale how much each issue impacted them by selecting the alternatives of Strongly Agree, Agree Somewhat, Neutral, Disagree Somewhat, and Strongly Disagree.

Table 5.9. Factors that deter respondents research data-sharing

		Strongly Agree		Somewhat Agree		Neutral		Somewhat Disagree		Strongly Disagree		Chi-Square p-value
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
There is no place to put the data	Q15.1	41	62.1%	6	9.1%	4	6.1%	6	9.1%	9	13.6%	< 0.001
Lack of incentives	Q15.2	39	60.0%	5	7.7%	6	9.2%	5	7.7%	10	15.4%	< 0.001
Lack of funding	Q15.3	32	48.5%	8	12.1%	9	13.6%	2	3.0%	15	22.7%	< 0.001
Lack of standards or guidelines for sharing data	Q15.4	40	60.6%	13	19.7%	4	6.1%	3	4.5%	6	9.1%	< 0.001
The data is not fully documented	Q15.5	18	27.7%	9	13.8%	10	15.4%	2	3.1%	26	40.0%	< 0.001
License agreements prohibit sharing data	Q15.6	31	47.7%	9	13.8%	9	13.8%	2	3.1%	14	21.5%	< 0.001
I would lose control over my data	Q15.7	20	30.3%	10	15.2%	7	10.6%	7	10.6%	22	33.3%	0.003
I have insufficient skills to make my data available to the public	Q15.8	16	24.6%	5	7.7%	7	10.8%	3	4.6%	34	52.3%	< 0.001
The data is in a format that is not widely readable	Q15.9	15	23.1%	4	6.2%	6	9.2%	7	10.8%	33	50.8%	< 0.001
My data may be misinterpreted by others	Q15.10	17	25.4%	8	11.9%	7	10.4%	5	7.5%	30	44.8%	< 0.001
The university owns the data I produce	Q15.11	32	47.8%	13	19.4%	11	16.4%	1	1.5%	10	14.9%	< 0.001
The funding agency owns the data	Q15.12	14	21.9%	11	17.2%	9	14.1%	3	4.7%	27	42.2%	< 0.001
Insufficient time	Q15.13	12	18.8%	6	9.4%	6	9.4%	4	6.3%	36	56.3%	< 0.001

Table 5.9 details the reasons that deterred respondents from sharing their collected data. Lack of a central location to deposit research data, with which 62.1% agreed strongly, 9.1% agreed somewhat, 9.1% disagreed somewhat, and 13.6% disagreed strongly; lack of incentives, with which 60.0% agreed strongly 7.7% agreed somewhat, 9.1% were neutral, and 7.7% disagreed somewhat; lack of standards, with which 60.6% agreed strongly, 19.7% agreed somewhat, 6.1% were neutral, 4.5% disagreed somewhat, and 9.1% disagreed strongly; lack of funding with 48.5% who agreed strongly, 12.1% agreed slightly; 13.6% were neutral, 3.0% disagreed somewhat, and 22.7% disagreed strongly; and license agreements prohibit data sharing with 47.7% who agreed strongly, 13.6% who agreed somewhat, 13.6% were neutral, 3.0% disagreed somewhat, and 21.5% disagreed strongly. From Table 5.12, these patterns were observed; some

statements showed significantly higher levels of agreement Q15.1 (62.1%); Q15.2 (60.0%); Q15.3 (48.5%); Q15.4 (60.6%), while six statements Q15.5 (40.0%); Q15.8 (52.3%); Q15.9 (50.8%); Q15.10 (44.8%); Q15.12 (42.2%) and Q15.13 (56.3%) indicate higher levels of disagreement. The contrast in these statements can be attributed to sub-themes that emerged, such as data ownership and control (Q15.7; Q15.9; Q15.10; Q15.11, Q15.12;) as well as central storage, incentives compliance (Q15.1; Q15.2; Q15.3; Q15.4; Q15.6). The significance of the differences is shown in highlights in Table 5.10.

Table 5.10.Significance levels of deterrents of research data-sharing

Statement	**Somewhat Agree**	**Neutral**	**Somewhat Disagree**
There is no place to put the data	**71.2**	6.1	22.7
Lack of incentives	**67.7**	9.2	23.1
Lack of funds	**60.6**	13.6	25.8
Absence of guidelines for sharing data	**80.3**	6.1	13.6
The data documentation is incomplete	41.5	15.4	**43.1**
License agreements prohibit sharing data	**61.5**	13.8	24.6
Loss of data control	**45.5**	10.6	43.9
Insufficient skills to make data available to the public	32.3	10.8	**56.9**

The data is in a format that is not widely understood	29.2	9.2	**61.5**
Data misinterpreted by others	37.3	10.4	**52.2**
The university owns the data I produce	**67.2**	16.4	16.4
The funding agency owns the data	39.1	14.1	**46.9**
Insufficient time	28.1	9.4	**62.5**

Findings indicate that researchers did not share their data since there was nowhere to put it. In contrast, several studies found that the main issue preventing researchers from sharing data was a lack of expertise in managing research data (Chawinga 2019; Yoon and Schultz 2017). The results also showed no standards or protocols for data sharing, which is why researchers did not share their data. A research conducted by Katabalwa *et al.* (2021) revealed that in Tanzania, the practice of sharing research data was relatively nascent. There were no mandates or incentives from governmental agencies to encourage research data sharing, resulting in the absence of formal research data management systems, open data repositories, and specific policies promoting research data sharing. The information was, therefore, not readily available. Therefore, it makes sense that readers would conclude a dearth of standards or norms for data management, given the absence of metadata. Researchers' lack of expertise may also be blamed for failing to record or assign metadata (Chawinga 2019).

5.6.5 Conditions for research data-sharing

The responders were given several options for indicating how much they would be encouraged to share the generated study data. Table 5.11 summarizes the findings.

Table 5.11. Patterns for conditions of research data-sharing

		Strongly Agree		Somewhat Agree		Neutral		Somewhat Disagree		Strongly Disagree		Chi-Square p-value
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
I would be willing to place at least some of my data into a central data repository with no restrictions	Q16.1	49	75.4%	8	12.3%	5	7.7%	2	3.1%	1	1.5%	< 0.001
I'd be willing to put all of my data in an open-access repository for data	Q16.2	45	69.2%	8	12.3%	7	10.8%	1	1.5%	4	6.2%	< 0.001
If I could impose restrictions on access, I would be more likely to make my data available	Q16.3	34	53.1%	13	20.3%	4	6.3%	2	3.1%	11	17.2%	< 0.001
I'd be willing to share data with numerous researchers who utilize it in various ways	Q16.4	47	72.3%	13	20.0%	4	6.2%	1	1.5%	0	0.0%	< 0.001
When other researchers use my data, they must cite me.	Q16.5	56	86.2%	5	7.7%	3	4.6%	0	0.0%	1	1.5%	< 0.001
It is acceptable to build new datasets using previously shared data	Q16.6	50	78.1%	9	14.1%	5	7.8%	0	0.0%	0	0.0%	< 0.001
Others	Q16.7	4	44.4%	2	22.2%	1	11.1%	0	0.0%	2	22.2%	0.550

Table 5.11 findings showed three main factors that could encourage respondents to share the data they generate: the significance of having their data cited when used, the appropriate creation of new sets of data derived from previous data transfer, and the willingness to give unrestricted access to a portion of their data in a single hub. The one component that attempted to understand data sharing if conditions for access are in place (I might be more inclined to make my data accessible if I could place access controls on it) failed to motivate researchers to share their data despite the almost marginal responses. The current study examined factors that facilitated or might have urged researchers to share their data. In contrast, previous research (Chawinga 2019: 185; Tenopir 2017: 45; Fecher *et al.* 2015: 21) suggested imposing

constraints on the data they supplied and having their data cited by reusers as the essential conditions that would encourage researchers to share data; the current study highlighted the researchers' desire of reusers citing their data when used; the appropriate generation of new datasets from shared data and the readiness to give some of their data to a central repository. The conditions of sharing depicted by the findings are placed upon citation, reusability, and data sharing across different disciplines. Table 5.12 shows in highlights the significance levels observed.

Table 5.12. Significance levels of conditions of research data sharing

Factors	**Somewhat Agree**	**Neutral**	**Somewhat Disagree**
I would be willing to place at least some of my data into a central data repository with no restrictions	**87.7**	7.7	4.6
I would be willing to place all my data into a central data repository with no restrictions	**81.5**	10.8	7.7
If I could impose restrictions on access, I would be more likely to make my data available	**73.4**	6.3	20.3
I'd be willing to share data with numerous researchers who utilize it in various ways	**92.3**	6.2	1.5
When other researchers use my data, they must cite me.	**93.8**	4.6	1.5
It is acceptable to build new datasets using previously shared data	**92.2**	7.8	0.0
Others	**66.7**	11.1	22.2

The results are at odds with those of a study by Bangani and Moyo (2019: 12), which presented that South African scholars favoured sharing their data with peers within their research environment rather than working with researchers at other universities due to expressed reservations about sharing data, including ethical considerations, data misuse, and resource shortages.

5.7 Research data reuse practices

In this section, the respondents were questioned about how frequently they utilized research data collected or produced by other academics or research institutes in their investigation activities to understand the respondents' perceptions about research data reuse and the application of the concept.

5.7.1 Research data reuse practices frequency.

This sub-section presented the frequency of use of data from the research created or produced by other scholars or institutions in the research activity fields of the respondents. This is shown in Figure 5.9.

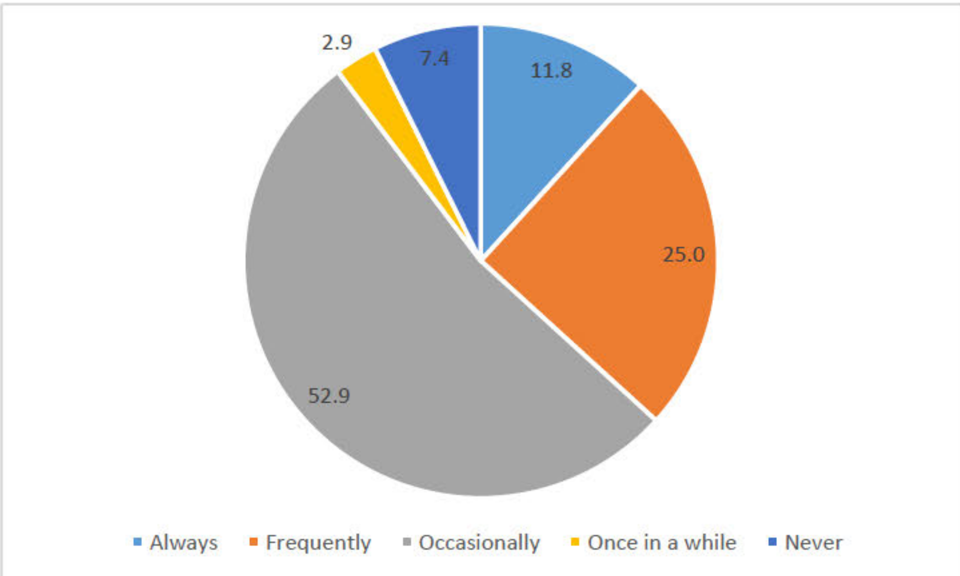

Figure 5.9. Respondents' research data reuse frequency

More than half of the respondents (52.9%) ($p < 0.001$) indicated they had occasionally engaged in research data reuse practices. The academic community is informed about the numerous advantages of data reuse by prior literature. According to Shakeri (2013: 87), reusing research data can reduce the expense and duplication of producing research data. Respondents risk losing on the countless profits of research data reuse. The findings of a study by Ng'eno (2018: 207) supported this finding. They showed that 88.7% of researchers in Kenyan research institutes profited from data exchange because it encouraged academic inquiry and discourse and reduced the cost of recreating data collection.

5.7.2 Factors that discouraged research data reuse

This question measured the extent to which respondents were less likely to use research data from other researchers or institutions in their research projects under various conditions. Respondents had to choose between the options of Strongly Agree, Agree Somewhat, Neutral, Disagree Somewhat, and Strongly Disagree to indicate on a Likert scale how much each matter influenced them.

Table 5.13. Factors that discouraged research data reuse

		Strongly Agree		Somewhat Agree		Neutral		Somewhat Disagree		Strongly Disagree		Chi-Square p-value
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Reusable data is hard to locate, find, or access	Q18.1	44	64.7 %	11	16.2 %	4	5.9%	5	7.4%	4	5.9%	< 0.001
Hard to integrate with my own data	Q18.2	34	50.7 %	6	9.0%	10	14.9 %	6	9.0%	11	16.4 %	< 0.001
Not trusting others' collection methods	Q18.3	21	32.3 %	10	15.4 %	7	10.8 %	5	7.7%	22	33.8 %	0.001
Data complexity raises the possibility of incorrect interpretation	Q18.4	17	25.8 %	11	16.7 %	5	7.6%	11	16.7 %	22	33.3 %	0.012
Lack of standardized formats	Q18.5	40	60.6 %	11	16.7 %	7	10.6 %	6	9.1%	2	3.0%	< 0.001
Inadequate metadata/data description	Q18.6	41	62.1 %	10	15.2 %	8	12.1 %	5	7.6%	2	3.0%	< 0.001
Data quality issues could lead to incorrect interpretations of the data	Q18.7	15	22.7 %	9	13.6 %	9	13.6 %	5	7.6%	28	42.4 %	< 0.001
The use of data may not always be what was intended.	Q18.8	18	26.9 %	16	23.9 %	7	10.4 %	4	6.0%	22	32.8 %	0.002
Legal/ethical restrictions	Q18.9	43	66.2 %	11	16.9 %	6	9.2%	2	3.1%	3	4.6%	< 0.001

Other	Q18.10	1	16.7%	0	0.0%	2	33.3%	0	0.0%	3	50.0%	0.607

The variables that kept respondents from adopting research data generated by other researchers are shown in detail in Figure 5.13. The critical criteria included the difficulty of finding, discovering, or accessing reusable data, with which 16.2% somewhat agreed and 64.7% strongly agreed. The second factor that discouraged respondents was legal or ethical limits, with which 66.2% agreed strongly, and 16.9% agreed somewhat. Inadequate metadata/data description came in at 62.1%, followed by a lack of standardized formats at 60.6%. According to the CCMF, many data aspects, including formats, collection techniques, descriptions, and data packaging and transfer protocols, should be thoroughly documented for data to be accessible by designated and possible reusers (Lyon *et al*. 2012: 131). Accordingly, it can be inferred from the results of the present study using the CCMF that it was challenging for researchers to use previously generated data that lacked criteria, which made it challenging to locate, unearth, or seize reusable data (Lyon *et al*. 2012: 131). Since they could not comprehend the circumstances surrounding the data's creation, its intended use, its creators, or its legal implications, proper metadata was proven to be crucial. According to these results, it is evident that some factors, such as the difficulty of integrating data, skepticism about the data collection techniques employed by other researchers, misinterpretation of intricate data, and the possibility that data might be used in ways other than what was intended, did not deter researchers from using data produced by other researchers.

There are significantly higher levels of agreement [Q18.1(64.7%); Q18.2(50.7%); Q18.5(60.6%); Q18.6(62.1%); Q18.9(66.2%)] associated with the factors that discourage research data reuse such as findability; lack of common standard formats and metadata. In his study, Chawing (2019) discovered that the non-standard assignment of metadata to data made it difficult to reuse research data. Similar to Brown *et al*. (2015: 85), Yoon (2015: 147), Woolfrey (2009: 79), and Musgrave (2003: 120), others have noted that standardized metadata causes interoperability issues. The lack of standardized metadata suggests that secondary users have trouble understanding and reusing primary investigators' data when it is shared. In that sense, the current study confirms earlier findings that comprehensive and consistent metadata can encourage researchers to become more interested in using data reuse in their research projects. The use and exchange of diverse data through numerous organizations, systems, and platforms are made more accessible by standardized metadata. The use of data requires trust. When incorporating researchers in the USA, Yoon (2015: 147) found that they were more inclined to use data generated by other researchers if they felt it came from trustworthy primary generators. Since metadata are the primary identifiers of the data's producers, their use is essential to demonstrating the data's uniqueness and reliability. As a result, using any data analysis tools makes it simpler for users to set interpretation and evaluation criteria. The bottom line is that data must be organized and deposited so researchers can obtain, share, and examine it (Tenopir *et al*. 2012: 70).

5.7.3 Research data preservation practices

This inquiry looked at the methods used to preserve research data. The necessity of preserving research data and the lifespan of research data were two specific topics that were looked into. Respondents were questioned about whether it was vital to keep research data safe. The results showed that 66 (98.5%) respondents believed retaining research data was important, whereas 1 (1.5%) disagreed. This is reflected in Table 5.14.

Table 5.14. Research data preservation frequency

	Frequency	Percent
Yes	67	98.5
No	1	1.5
Total	68	100.0

Significantly more respondents indicated that they believed it was important to preserve their research data ($p < 0.001$).

The researcher asked respondents how long their stored data would be useful and accessible.

Table 5.15. Respondents' research data preservation life-span

	Frequency	Percent

Indefinitely	4	5.9
10 – 20 years	4	5.9
5–10 years	27	39.7
3–5 years	29	42.6
1-2 years	3	4.4
Not sure	1	1.5
Total	68	100.0

Table 5.15's data reveals that 5.9% respondents thought their data would be useful indefinitely; similarly, 5.9% believed that their research data would be useful in 10-20 years, 39.7% thought it would be useful in five to ten years, 42.6% thought it would be useful in three to five years, 4.4% thought it would be useful in 1-2 years, and 1.5% was unsure. According to an analysis of these findings, most respondents believed their data would be valuable for three to ten years.

Significantly more of the respondents (81.8%) believed that data would remain valuable for 3 to 10 years ($p < 0.001$). Table 5.15 highlights the significance of a central platform for depositing research data and recommendations for data preservation and retention policies. According to Chawinga (2019), one of the universities understudied lacked a centralized data management system. Unfortunately, there were no adequate backups for most of the data stored on flash disks, external hard drives, computers, or laptops in departmental laboratories. Since computers and laptops frequently crash or are lost, most researchers lose their data. These results are further supported by those found in section 5.5.3, which revealed that respondents' most popular digital storage options were personal computers, office computers, emails, external hard drives, freely downloadable software or services like Google Drive, and flash disks.

5.8 Design for a faculty research data repository platform based on research evidence

This section looked at essential elements and features that can guide the creation of a research data repository depending on the respondents' needs.

Respondents were asked to characterize the assistance they thought their faculty should offer to help with research data management tasks in a follow-up question. Participants were asked to respond to the issue using a Likert scale, choosing from the options of Agree Strongly, Agree Somewhat, Neutral, Disagree Somewhat, and Disagree Strongly to express how much they desired their faculty to provide each type of support.

Table 5.16. Infrastructure support to be provided by the faculty

		Strongly Agree		Somewhat Agree		Neutral		Somewhat Disagree		Strongly Disagree		Chi-Square p-value
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Should create a procedure for data management throughout the project's life (short term – 5 years or less)	Q23.1	50	82.0%	7	11.5%	3	4.9%	1	1.6%	0	0.0%	< 0.001
After the project, a process for managing data should be established (long-term beyond five years).	Q23.2	52	86.7%	3	5.0%	4	6.7%	1	1.7%	0	0.0%	< 0.001
Should set up the necessary equipment and technical assistance for data management throughout the project's lifecycle (short term – 5 years or less)	Q23.3	47	79.7%	7	11.9%	4	6.8%	1	1.7%	0	0.0%	< 0.001
Establish the required hardware and software for data	Q23.4	48	80.0%	8	13.3%	4	6.7%	0	0.0%	0	0.0%	< 0.001

management, extending beyond the project's duration (long-term beyond five years).												
Should allocate the funds required to support data management throughout a research project (short-term, five years or less)	Q23.5	32	54.2%	10	16.9%	0	0.0%	2	3.4%	15	25.4%	< 0.001

Table 5.16's data reveals that 82.0% participants strongly agreed, 11.5% participants somewhat agreed, and 4.9% participants had mixed opinions (neutral) about the necessity of establishing a process for data management for five years or less; 86.7% of participants strongly agreed, 4.9% participants agreed somewhat agreed, 6.7% participants were neutral, 1.7% participants disagreed somewhat on the need to establish a process for managing data beyond a period of five years. The third factor addressed establishing tools and technical assistance required for data management throughout the project's life (short-term – 5 years or less), and participants responses showed that 79.7% agreed strongly, 11.9% agreed somewhat, and 6.8% were neutral; while for establishing necessary tools and technical support for data management data beyond the life of the project(long-term - beyond five years), the respondents 80% agreed strongly, 13.3% agreed somewhat, and 6.7% were neutral. In terms of providing the funds required to boost the management of data throughout a research project (short-term -5 years or less), 54.2% strongly agreed, 16.9% somewhat agreed, 3.4% disagreed somewhat, and 25.4% strongly disagreed.

Significantly higher levels of agreement have been observed with question statements [Q23.1(82.0%); Q23.2(86.7%); Q23.3(79.7%); Q23.4(80.0%); Q23.5(54.2%)] associated with offering each of the five types of support services for research data management. This aligns with various studies that commend the establishment of research data management platforms for long-term preservation. The infrastructure for research data goes far beyond the less dependable data storage devices like laptops and their accessories like flash drives and external hard drives. An integrated system combining hardware, software, and human resources defines a reliable and robust data infrastructure (Kabanda *et al.* 2023: 9). According to Brown *et al.* (2015: 86), a system for data management infrastructure will make it easier to manage data generally and to preserve and access it both temporarily and permanently. According to Chawinga's study (2019: 194), the lack of a research data infrastructure results in not having a formal system in place for collecting data from primary creators or data repositories and preparing it for prolonged incubation and tagging it with the appropriate metadata to make it easier to access and reuse. To successfully manage both short-term and long-term data, Schöpfel *et al.* (2018: 248) and Schumacher and Vande Creek (2015: 99) claim that universities must provide networked storage, non-networked devices, and networks that colleges or universities manage. Researchers are discouraged from participating in RDM activities due to a lack of such data infrastructure, making the data they produce not formally preserved. The CCMF recommends investing in computer-based extensive petabyte-scale research data processing for long-term storage, retention, access, and reuse (Lyon *et al*. 2012: 133).

5.9 Cross-tabulations on research data attributes

The section on cross-tabulations explains some of the patterns that have been identified to have significantly influenced the views and understanding of some of the factors of the study. The section dealing with data attributes (types of data employed by respondents) was used to present cross-tabulations. Pearson’s Chi-square statistical test was employed to help determine the results' significance. The likelihood of an observed difference between the study's categorical data sets can be determined using the Pearson chi-square test. The null hypothesis states that any difference between the sets of categorical data is due to experimental or sampling error and that there is no discernible difference between them (Pearson 1901: 559). Table 5.17 summarises the results of the chi-square tests to the statement that sought to understand the digital facilities used by respondents as the location for their research data sets.

Table 5.17. Cross-tabulations of databases used by respondents

Gender	Total

			Male	Female	
Digital databases (e.g., surveys, census data, government statistics, etc.)	Yes	Count	5	17	22
		% within Digital databases (e.g., surveys, census data, government statistics, etc.)	22.7%	77.3%	100.0%
		% within Gender	17.9%	42.5%	32.4%
		% of Total	7.4%	25.0%	32.4%
	No	Count	23	23	46
		% within Digital databases (e.g., surveys, census data, government statistics, etc.)	50.0%	50.0%	100.0%
		% within Gender	82.1%	57.5%	67.6%
		% of Total	33.8%	33.8%	67.6%
Total		Count	28	40	68
		% within Digital databases (e.g., surveys, census data, government statistics, etc.)	41.2%	58.8%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	41.2%	58.8%	100.0%

The p-value between "Digital databases (e.g., surveys, census data, government statistics, etc.)" and "Gender" is 0.039. It is noted that significantly more females said yes (42.5%) while significantly more males indicated no (82.1%). This means that there is a significant relationship between the variables (digital databases and gender); the gender of the respondents played a substantial role in how respondents viewed and understood digital facilities.

Table 5.18. Cross-tabulations of devices and gender

			Gender		Total
			Male	Female	
Office computers	Yes	Count	17	14	31
		% within Office computers	54.8%	45.2%	100.0%

		% within Gender	60.7%	35.0%	45.6%
		% of Total	25.0%	20.6%	45.6%
	No	Count	11	26	37
		% within Office computers	29.7%	70.3%	100.0%
		% within Gender	39.3%	65.0%	54.4%
		% of Total	16.2%	38.2%	54.4%
Total		Count	28	40	68
		% within Office computers	41.2%	58.8%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	41.2%	58.8%	100.0%

The p-value between "the frequency of research data back up on office computers" and "Gender" is 0.049. This indicates a strong correlation between the variables (office computer and gender); gender did play an important influence on participants' perceptions back up the frequency of in-office computers.

5.10 Correlations of research data-sharing variables

The ordinal data were also subjected to bivariate correlation. In statistics, the correlation coefficient gauges how strongly two variables interact linearly. This method is frequently used to describe direct connections without stating cause and effect. (Tu and Shen 2007: 1287). The researcher used factors that deter students from sharing research data to showcase correlations, indicating the following patterns.

For example, the correlation value between "The data is not fully documented" and "The data is in a format that is not widely readable" is 0.597. This is a directly related proportionality. Respondents indicated that the more documents are not fully documented, the more likely the document would be in an unreadable format, and vice versa. Positive values indicate a directly proportional relationship between the variables. An inverse relationship is characterized by negative values or one where the variables affect one another in the opposite direction. For example, the correlation value between "There is no place to put the data" and "Insufficient time" is -0.291. The more there is no place to put the data, the less time there is to do so.

The Chi-square test ($p < 0.001$) shows a significant correlation value between "There is no place to put the data" and "I'd be willing to put all of my data in an open-access repository for data." The correlation coefficient is 0.628. Respondents indicated that the more likely there is a place to put data, the more likely they would be willing to place their research data without restrictions. The findings also reveal that if "The data is in a format that is not widely readable," respondents are less likely willing to place all their data into a central data repository with no restrictions. The correlation coefficient is -0. 321.

The results further reveal a significant correlation value between "the lack of standards or guidelines for sharing data" and "establishing a process for managing data during the life of the project (short-term – 5 years or less)." The correlation efficiency is 0.406. The more likely processes are established for managing research data, the more likely it will address the lack of standards or guidelines for sharing research data. An inverse relationship is also observed in the variable statements about "establish a process for managing data during the life of the project (short-term – 5 years or less)" and "others may misinterpret my data." The correlation coefficient is -0.288.

5.11 Factor Analysis of variables

In this study, factor analysis was employed. The statistical technique known as factor analysis' primary goal is to reduce the amount of data. Factor analysis is often used in the survey method when a researcher is interested in representing numerous research questions with a few made-up variables (Du 2009: 258). For the factor analysis, only the Likert scale items are utilized. Some components are broken down into smaller components. In this case, factor analysis was done in questions Q15, Q16, Q18, and Q23.

A summary table (Table 5.22) containing the Kaiser-Meyer-Olkin (KMO) and Bartlett's Test results is presented before the matrix tables below. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy statistic indicates the percentage of variance in the variables that underlying factors may contribute to. High values suggest that factor analysis may benefit the data (values near 1.0). The factor analysis findings won't be useful if the value is less than 0.5. Bartlett's test of homogeneity of variance is used to verify the assertion that the correlation matrix is an unbiased estimator, which would suggest that the variables have no connection and, as a result, are inappropriate for categorical attributes (Du 2009: 258). If the significance level for the data is low, factor analysis may be useful (less than 0.05). Table 5.19 includes the **KMO and Bartlett's Tests** and lists two tests that show whether data are suitable for structure detection.

Table 5.19. KMO and Bartlett's Test

	Section	Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	Bartlett's Test of Sphericity		
			Approx. Chi-Square	df	Sig.
Q15	Factors that deter students from sharing research data	0.727	402.152	78	< 0.001
Q16	Conditions for research data sharing	0.682	234.674	15	< 0.001
Q18	Factors that discouraged research data reuse	0.738	334.212	36	< 0.001
Q23	A Design for a faculty research data repository platform based on research evidence	0.598	161.496	10	< 0.001

The requirements for factor analysis are all met.

Bartlett's Test of Sphericity significance value and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy value should be greater than or equal to 0.500.

Rotated Component Matrix

Principal component analysis and Varimax with Kaiser Normalization were used as the rotation and extraction methods, respectively. Very few factors place heavy demands on each attribute using this orthogonal rotation technique. It makes the factors' interpretation simpler.

Table 5.20. Rotated Component Matrix loading

Q15	Component		
	1	2	3
There is no place to put the data	-0.199	0.851	-0.077
Lack of incentives	0.171	0.780	0.082
Lack of funding	0.420	0.676	0.172
Lack of standards or guidelines for sharing data	0.011	0.758	0.042
The data is not fully documented	0.713	0.317	0.092
License agreements prohibit sharing data	0.373	0.331	0.641
I would lose control over my data	0.815	0.097	0.082
I have insufficient skills to make my data available to the public	0.884	-0.018	0.033
The data is in a format that is not widely readable	0.776	0.180	0.127
My data may be misinterpreted by others	0.901	-0.049	0.025
The university owns the data I produce	-0.324	-0.080	0.541
The funding agency owns the data	0.164	0.040	0.801
Insufficient time	0.725	-0.042	-0.062

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Table 5.21. Variables reduction and normalization

Q16	Component	
	1	2
I would be willing to place at least some of my data into a central data repository with no restrictions	0.857	-0.186
I would be willing to place all my data into a central data repository with no restrictions	0.830	-0.201
I would be more likely to make my data available if I could place conditions on access	0.020	0.948
I would be willing to share data across a broad group of researchers who use data in different ways	0.818	0.228
It is important that my data are cited when used by other researchers.	0.804	0.130
It is appropriate to create new datasets from shared data	0.898	0.105

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Table 5.22. Data reduction and normalization

Q18	Component	
	1	2
Difficult to find, discover, or access reusable data	0.788	-0.111
Hard to integrate with my own data	0.750	0.219
Not trusting others' collection methods	0.220	0.734
Data may be misinterpreted due to complexity of the data	0.118	0.878
Lack of common or standard formats	0.864	-0.015
In adequate metadata/data description	0.903	0.021
Data may be misinterpreted due to poor quality of the data	-0.080	0.802

Data may be used in other ways than intended	-0.028	0.910
Legal/ethical restrictions	0.766	0.172

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Table 5.23. Variables for data reduction and normalization

Q23	Component	
	1	2
Should establish a process for managing data during the life of the project (short term – 5 years or less)	0.886	0.166
Should establish a process for managing data beyond the life of the project (long term beyond five years).	0.784	-0.224
Should establish necessary tools and technical support for data management during the life of the project (short term – 5 years or less)	0.888	0.123
Should establish necessary tools and technical support for data management data beyond the life of the project(long term beyond five years).	0.824	-0.238
Should establish necessary funds to support data management during the life of a research project (short term, five years or less)	-0.025	0.950

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Factor analysis loading revealed inter-correlations between variables, as shown in Tables 5.21, 5.22, and 5.23. Questions' items with similar loadings implied measurement along the same axis. A review of items with loading values of at least 0.5 is successfully analyzed along the numerous components. It is noted that variables constituted the various questions loaded along 2 or 3 components (sub-themes) which are colour coded. This indicates that respondents reported different trends within the section. For example, trends identified in component 1 were associated with challenges in research data management; component 2 trends central storage, incentives, and compliance; and component 3 exhibited trends around research data ownership and control.

5.12 Structural Equation Model

This investigation also conducted the analysis using structural equation modelling (SEM). A group of statistical techniques known as SEM are used to quantify and analyze the relationships between latent and observed variables (Beran and Vialato 2010: 267). Similar to, but superior to regression analyses, it examines linear causal relationships between variables while considering measurement error (Beran and Vialato 2010: 267). SEM is widely used to evaluate research hypotheses in the social sciences, such as business, sociology, psychology, and education (Mueller, 1996; Schumacker and Lomax, 2004). According to Thompson (2000: 271), it would be challenging to find current issues related to social scientific articles where some SEM implementations were not disclosed.

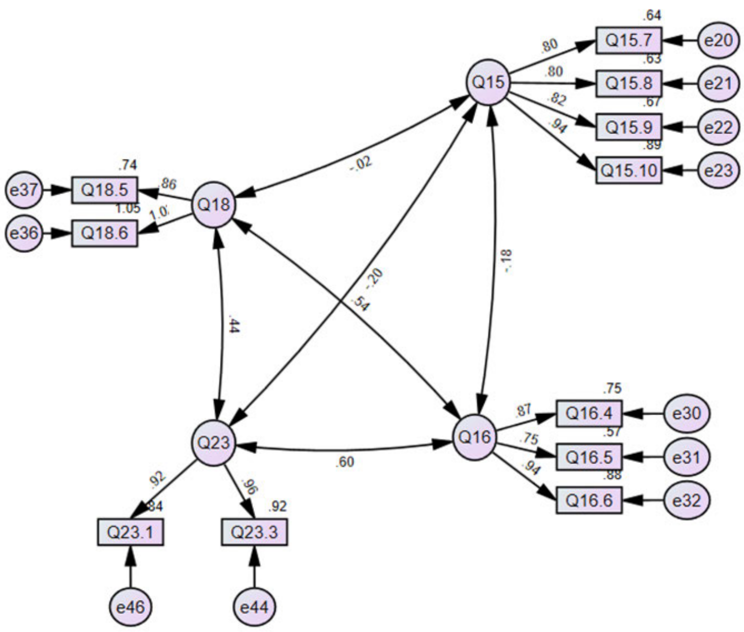

Figure 5.10.SEM path diagram

The data used to create the path model above came from a survey of masters and doctoral students in the faculty of Accounting and Informatics between 2015 and 2020. The path diagram in Figure 5.10 shows the measurement of the theoretical constructs that deter students from sharing research data (Q15), the conditions for sharing research data (Q16), the factors that discouraged the reuse of research data (Q18), and the factors that direct the development of a repository for research data (Q23). Three items measured the conditions for research data sharing; two measured the factors that discouraged research data reuse; four measured the factors that discouraged students from sharing research data; and two items measured the factors that guided the creation of a research data repository. The measured constructs, namely, Q15, Q16, and Q18, were aligned with the research objective RO 3, which sought to "establish the usefulness of the faculty research data repositories." In contrast, Q23 was aligned with RO 4, which sought to "design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence." The dimensions are coded as under reliability.

Sometimes, researchers need to confirm a path model to validate an instrument using multiple samples. This Chi-square analyzes the just-identified (full, saturated) model and tests the null hypothesis that the overidentified (reduced) model fits the data. Each variable in the newly discovered model has a direct path to every other variable without passing through an intermediary variable (Chau 1997: 310; Hair *et al.* 2016: 212). In such a model, the fit will always be perfect, so the Chi-square will always be equal to zero. The likelihood should not matter much. The chi-square p-value for this model is > 0.050 ($p = 0.11$). The prerequisite for a good model fit is satisfied (Wang *et al.* 2005: 7). Table 5.24 used the Maximum Likelihood Estimates to determine the probability distribution parameters.

5.12.1 Maximum Likelihood Estimates

Myung (2003: 95) asserts that the concept of maximum likelihood estimation (MLE), first put forth by R.A. Fisher in the 1920s, stipulates that the "desired probability distribution is the one that makes the observed data most likely," which necessitates the search for the value of the parameter vector that maximizes the likelihood function. Tables 5.24 and 5.25 demonstrate the probability distribution corresponding to this MLE estimate. According to the MLE principle, these factors are likely to have produced the observed data coefficient above the suggested value of 0.600.

Table 5.24. Regression Weights

			Estimate	S.E.	C.R.	P	Label
Q15.7	<---	Q15	1.000				
Q15.8	<---	Q15	1.015	.143	7.114	***	par_1
Q15.9	<---	Q15	1.019	.138	7.408	***	par_2
Q15.10	<---	Q15	1.197	.138	8.698	***	par_3
Q16.4	<---	Q16	1.000				
Q16.5	<---	Q16	.876	.122	7.181	***	par_4
Q16.6	<---	Q16	.976	.101	9.643	***	par_5
Q18.6	<---	Q18	1.000				
Q18.5	<---	Q18	.854	.107	7.989	***	par_6
Q23.3	<---	Q23	1.107	.124	8.916	***	par_7
Q23.1	<---	Q23	1.000				

The variables loaded strongly along their various factors (significant p-values indicated by *** $p < 0.001$). These verify the Exploratory Factor Analysis (EFA) obtained under factor analysis.

Table 5.25. Standardized Regression Weights

			Estimate
Q15.7	<---	Q15	.797
Q15.8	<---	Q15	.796
Q15.9	<---	Q15	.819

	Estimate
Q15.10 <--- Q15	.943
Q16.4 <--- Q16	.867
Q16.5 <--- Q16	.752
Q16.6 <--- Q16	.940
Q18.6 <--- Q18	1.000
Q18.5 <--- Q18	.860
Q23.3 <--- Q23	.959
Q23.1 <--- Q23	.918

Maximum Likelihood (ML) methods are used to determine the parameters. This iterative process seeks to maximize the likelihood that the obtained values of the independent variables will be correctly predicted. The coefficients were all higher than the recommended value of 0.600.

5.12.2 Model Fit Summary

The key metrics used to choose the ideal starting point for the final model are gathered in the Model Fit Summary. These model fit indices track differences between actual correlation/covariance matrices and those implied by the model. The differences between observed and model-implied data are typically represented by model fit indices (Stanley and Edwards 2016: 978). The suggested acceptable value for relative chi-square, to lessen reliance on sample size, the Minimum Discrepancy Function by Degrees of Freedom divided (CMIN/DF) should not exceed 5. The Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Nominal Fit Index (NFI), and Incremental Fit Index (IFI) all have a cut-off range between zero and one. A root mean square error of approximation (RMSEA) value of less than or equal to 0.05 denotes a good model.

Model Fit Summary

Table 5.26. CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	39	48.901	38	.111	1.287
Saturated model	77	.000	0		
Independence model	11	563.793	66	.000	8.542

As indicated in Table 5.26, the result of the default model minimum was achieved with the Chi-square = 48.901, Degrees of freedom = 38, and Probability level = .111. CMIN is a Chi-square statistic comparing the tested and the independent models to the saturated models (Beran and Vialato 2010: 269). The relative chi-square ratio, CMIN/DF, measures how much the model's ability to fit the data has been compromised by eliminating one or more paths. CMIN/DF is below the acceptable level of 5. (1.287). This satisfies the CMIN requirement.

Table 5.27. Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.913	.849	.979	.962	.978
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Instead of comparing the model to the saturated model, these goodness of fit indices compare it to the independence model. The difference between the chi-squares for the two models, divided by the chi-square for the independence model, is known as the Normed Fit Index (NFI) (Beran and Vialato 2010: 269). The NFI for these data is 0.913, which is higher than the suggested value of 0.90 for a good fit. Even with small samples, the Comparative Fit Index (CFI), which employs a non-central chi-square, uses a similar methodology. The NFI ranges from 0 to 1, and a fit of 0.90 is considered good. The CFI value is 0.978, as indicated in Table 5.27, indicating a good fit.

Table 5.28. The Root Mean Square Error of Approximation

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.065	.000	.114	.305
Independence model	.336	.310	.361	.000

When compared to the saturated model, the Root Mean Square Error of Approximation (RMSEA) calculates the lack of fit (Beran and Vialato 2010: 269). A good fit is an RMSEA of 0.050 or less between .05 and 08. The lower and upper limits of a 90% confidence interval are designated LO 90 and HI 90, respectively. The PCLOSE value (0.305) is insignificant, and the model provides an adequate fit. This criterion has been satisfied (recommended > 0.050).

5.12.3 Regression Analysis

Regression analysis is one of the most popular analysis methods in research to examine the connections between dependent and independent variables (Sarstedt and Mooi 2019: 210). Sarstedt and Mooi (2019: 210) further assert that the main advantages of using regression analysis are that it enables us to estimate the relative strength of various independent variables' effects on a dependent variable, determine whether one independent variable or a group of independent variables have a significant relationship with a dependent variable, and make predictions. The level of significance relates to the strength of the connections. The correlations tested are shown in Tables 5.29 and 5.30.

Table 5.29. Co-variances

	Estimate	S.E.	C.R.	P	Label
Q15 <--> Q18	-.037	.192	-.193	.847	par_8
Q15 <--> Q16	-.143	.107	-1.335	.182	par_9
Q23 <--> Q15	-.153	.108	-1.412	.158	par_10
Q16 <--> Q18	.360	.099	3.659	***	par_11
Q23 <--> Q18	.289	.094	3.073	.002	par_12

	Estimate	S.E.	C.R.	P	Label
Q23 <--> Q16	.202	.056	3.590	***	par_13

Table 5.30. Correlations

	Estimate
Q15 <--> Q18	-.024
Q15 <--> Q16	-.183
Q23 <--> Q15	-.198
Q16 <--> Q18	**.542**
Q23 <--> Q18	**.439**
Q23 <--> Q16	**.603**

Since correlation is a standardized form of co-variance and both measure the strength and direction of the relationship, it follows that if there is a significant co-variance between two constructs, there should also be a significant correlation between them (Beran and Vialato 2010: 271). Table 5.30's three highlighted relationships are significant (p 0.05). The findings show that the latent variables have a strong, directly proportional relationship with each positive r estimate. As an illustration, a rise in Q16 causes a rise in Q18 (r = 0.542), and vice versa. The model did not include the low-loading factors or statements. High factor loadings were seen when the coefficients for each latent variable were examined. The diagram in Figure 5.10 also consists of a reflection of the path coefficients. Between the latent variables, there are positive coefficients for every proportional relationship. It is also expected that the structural relationships might not have fitted precisely because this was a newly developed construct. All indices are met, and the model fits well, but it is advised to be revised in subsequent studies so that the measured variables serve as the latent variables to enhance factor loadings.

5.13 Appraisal of the Chapter

Chapter five presented a quantitative analysis of the survey conducted among masters and doctoral students within the Faculty of Accounting and Informatics from 2015 to 2020. The analysis utilized various methods, including graphs, tables, frequencies, and percentages, all aligned with the study's research objectives. These visual representations effectively communicated the patterns observed in research data management within the faculty.

The results indicated that research data management was primarily seen as an individual responsibility rather than a structured, collective effort within the faculty. Notably, the lack of investment in research data management emerged as a significant challenge. The absence of centralized systems for managing research data was highlighted, with researchers relying primarily on personal devices like laptops, emails, and external hard drives.

Despite these challenges, participants acknowledged the benefits of effective research data management, such as scientific advancements, cost-effectiveness in data reuse, and enhanced convenience for data reusers. A key finding was the desire among researchers to share their data without restrictions, illustrating a positive attitude towards data sharing within the academic community.

However, the study revealed a critical obstacle to data sharing and reuse: the lack of appropriate platforms that support these practices. Such platforms' absence hindered researchers from sharing and reusing data. Importantly, researchers were willing to share their data without limitations, underscoring the importance of establishing suitable data-sharing mechanisms to facilitate collaboration and knowledge advancement in the academic community.

CHAPTER SIX: PRESENTATION OF QUALITATIVE RESULTS AND DISCUSSION

6.1 Introduction

The study gathered qualitative information by conducting a meta-analysis of research repositories at HEIs and in-depth interviews with research supervisors from the Faculty of Accounting, as mentioned in Chapter Four. The interview guide attached in the annexure (See Annexure I) was the basis of the researcher interviews. Creswell (2018: 109) describes data presentation and analysis as a sequence of procedures by researchers to identify significant themes from the study, followed by a discussion of the results. To ensure research continuity, the discussions should "establish relationships within the collected data, partially overlapping analysis" (Kothari 2004: 344). It should also compare study findings to those of other related studies that have already been conducted (Hess 2004: 1172). The researcher is advised to focus on the research objectives and the results throughout this process (Hess 2004: 1172).

The study's objective was to examine how research data are managed within the faculty environment of a university of technology to support researchers by designing a prototype for a research data repository appropriate for such a setting. Data was gathered from the faculty of Accounting and Informatics supervisors through online interviews conducted via ZOOM and MS TEAMS. According to the research questions, the data were coded, categorized into themes, and presented as figures, narratives, and verbal descriptions. The process by which numerous words within a text are grouped into fewer categories or themes is known as thematic content analysis (Weber 1990: 15). Using NVIVO software, the researcher could group and sort data and look at relationships. A meta-analysis of the most significant papers on research repositories was incorporated into the analyses and interpretations of results to provide support and produce a richer, more in-depth understanding of the issue. In addition to the participants' demographics, the following topics are presented in this chapter as findings based on the study's research questions:

- Analyze the most pertinent studies on research repositories at Higher Education Institutions using machine learning.
- Ascertain the management of the existing research data
- Establish the usefulness of the faculty research data repositories

- Design a novel digital prototype for a faculty research data repository platform based on research evidence

For a visual representation of the chapter's contents, discussion points, and findings, the researcher employed the explanatory sequential mixed method study diagram illustrated in Figure 6.1.

Explanatory Sequential Mixed Method Study

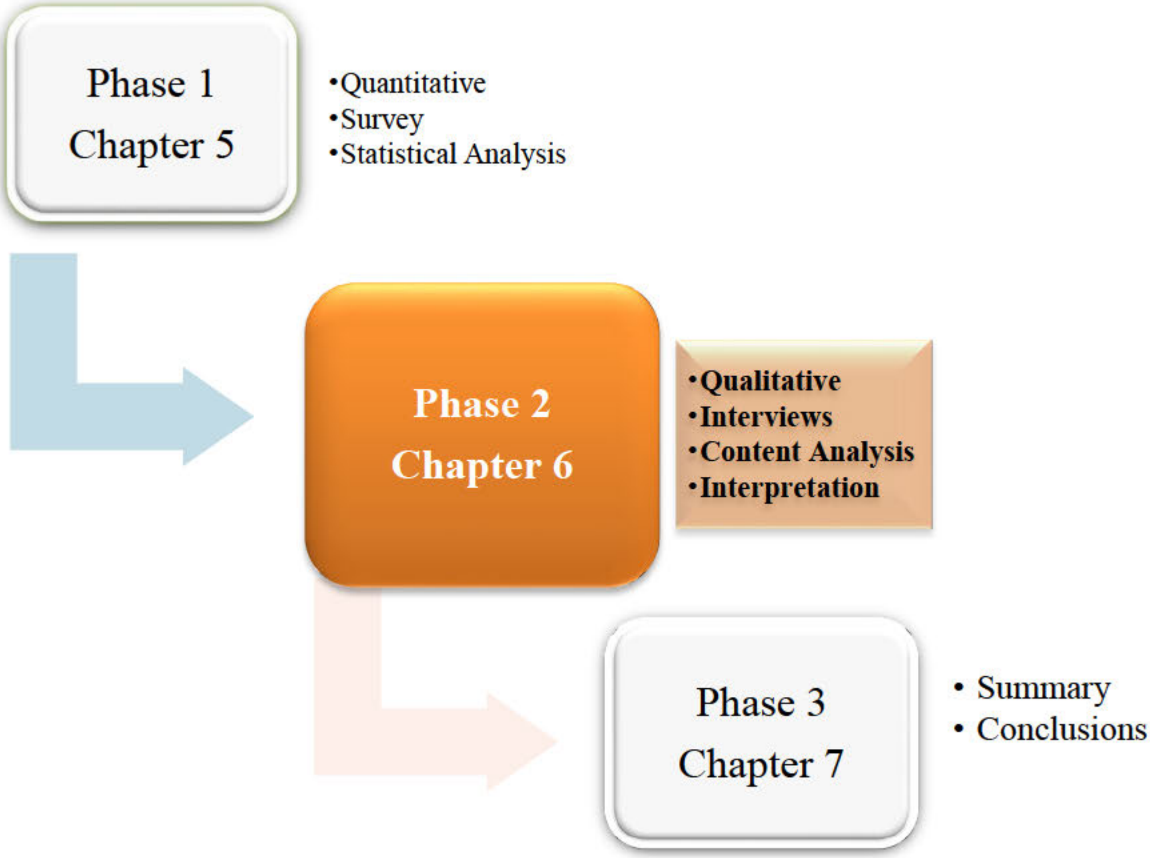

Figure 6.1. An adaptation of Creswell Explanatory Sequential Mixed Method Study

Source: Creswell and Plano Clark (2018)

6.2 A Meta-Analysis of the research repositories at HEI using machine learning

A comprehensive review of the most significant papers on research repositories was conducted to understand the motivations for developing HEI research repositories. This review was to discover the benefits of faculty research repositories, their services, and how they are used. This section summarizes the findings of a thorough analysis of research repository practices in global HEIs. It offers helpful information on the origins, uses, benefits, and acceptance of research repository practices by departments and faculties in HEIs.The arguments are then summarized based on the assessment parameters provided in Chapter Three, section 3.1. A thorough search was conducted on pertinent articles published in databases between January 2015 and March 2021. A qualitative synthesis was performed on the eleven relevant full-text articles after 24 full-text articles were excluded since they did not cover the complete scope of the review (see Chapter Four, section 4.3 for an in-depth approach).

6.2.1 Latent Dirichlet Analysis

The semantic association between the eleven research repositories in HEIs' chosen articles was visualized using the Latent Dirichlet Allocation (LDA), a machine learning algorithm. Using LDA, subjects (specific terms or phrases) were found to suggest the recurring themes in the text corpus (eleven publications). According to Saib *et al.* (2022: 15), LDA assumes that papers are a mixture of subjects. It was used to visualize these topics, which were prevalent in the eleven publications chosen for the study. The clustering algorithm of unstructured document categorization, the foundation for the LDA studies, revealed four prominent subjects with keywords from the sample of eleven articles. The following topics have been turned into themes: Topic 0 (“record management”), Topic one (“research repository”), Topic two (“research publication data”), and Topic three (“participant data”), as shown in Table 6.1.

Table 6.1. Topics and keywords extracted by latent Dirichlet allocation (LDA)

Topic	Theme	Keywords

0	Record Management	Record, repository, faculty, project, management, study, student, system, user, create
1	Research Repository	Research, simulation, repository, exposure, nursing, university, include, study, publication, topic.
2	Research publication data	Publication, research, datum, rank, collection, faculty, project, digital, file, index
3	Participant data	Datum, ping, participant, base distortion, cortical, analysis, available, include measure.

Figure 6.2 demonstrates the topic phrases and their weights from the corpus output of the model using article titles to streamline the analysis.

```
[(0,
  '0.027*"record" + 0.014*"faculty" + 0.013*"management" + 0.011*"repository" '
  '+ 0.011*"study" + 0.007*"user" + 0.007*"system" + 0.007*"respondent" + '
  '0.007*"electronic" + 0.006*"policy"'),
 (1,
  '0.024*"research" + 0.020*"repository" + 0.019*"project" + '
  '0.013*"simulation" + 0.013*"student" + 0.009*"exposure" + 0.009*"nursing" + '
  '0.009*"include" + 0.008*"faculty" + 0.008*"university"'),
 (2,
  '0.015*"publication" + 0.013*"research" + 0.012*"datum" + 0.011*"rank" + '
  '0.010*"collection" + 0.010*"faculty" + 0.010*"project" + 0.009*"digital" + '
  '0.009*"index" + 0.008*"file"'),
 (3,
  '0.018*"datum" + 0.011*"ping" + 0.009*"participant" + 0.006*"base" + '
  '0.005*"distortion" + 0.005*"cortical" + 0.005*"include" + 0.005*"analysis" '
  '+ 0.005*"available" + 0.004*"age"')]
```

Figure 6.2. Keyword and weight from the corpus

The top keywords for each topic are represented as word clouds in the picture below. The magnitude of the words gives a hint as to how likely it is that the word in the corpus belongs to that issue. Topics are categorized by colour; blue represents topic 0; orange is used to signify Topic 1. Topic 2 is depicted in green, whereas Topic 3 is shown in red.

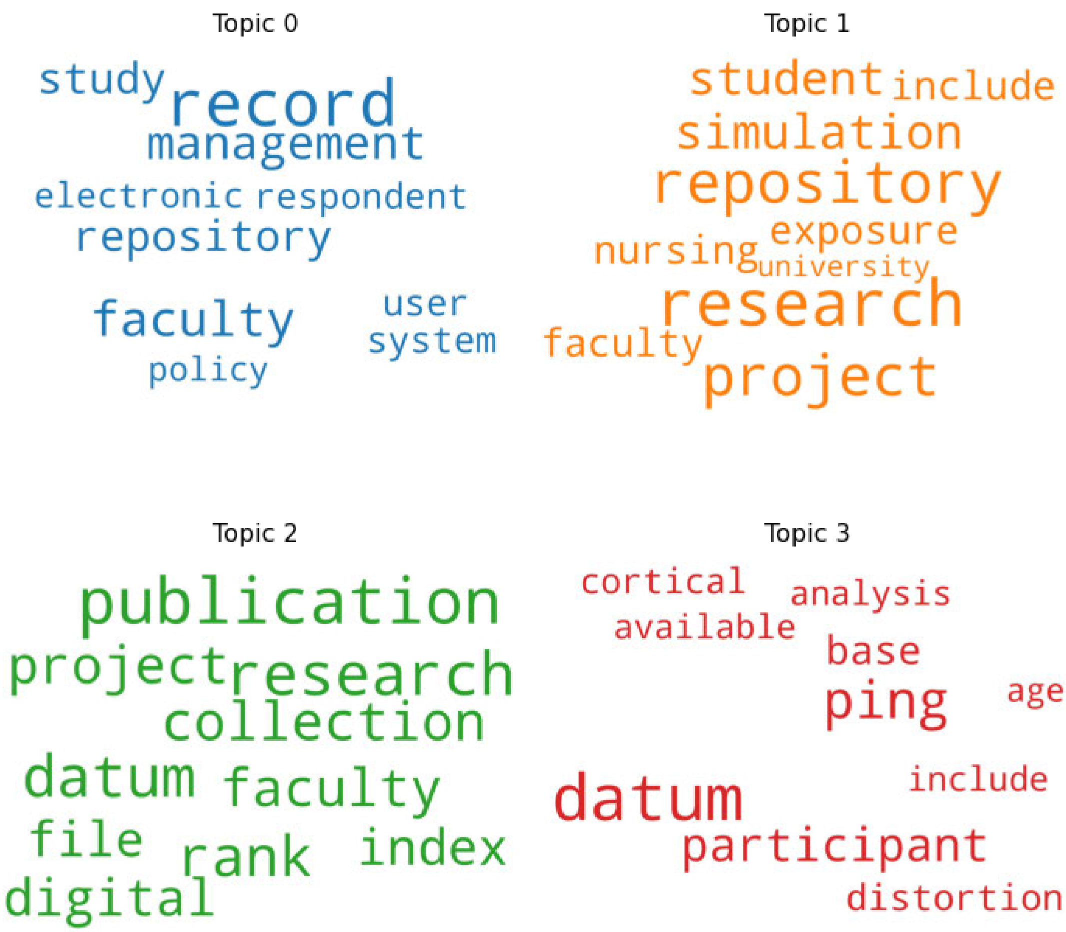

Figure 6.3. Word Clouds of Top keywords in each Topic

The graph below demonstrates that the terms "research," "repository," and "project" dominated the corpus of the eleven selected articles, whereas "record," "faculty," and "management" were the terms most frequently used in Topic 0. Four subjects (concepts) from the corpus of data are displayed on the inter-topic distance map. The figure shows the overlap between topics one and three: "Record Management," "Research publishing data," and "Research Repository versus Participant data."

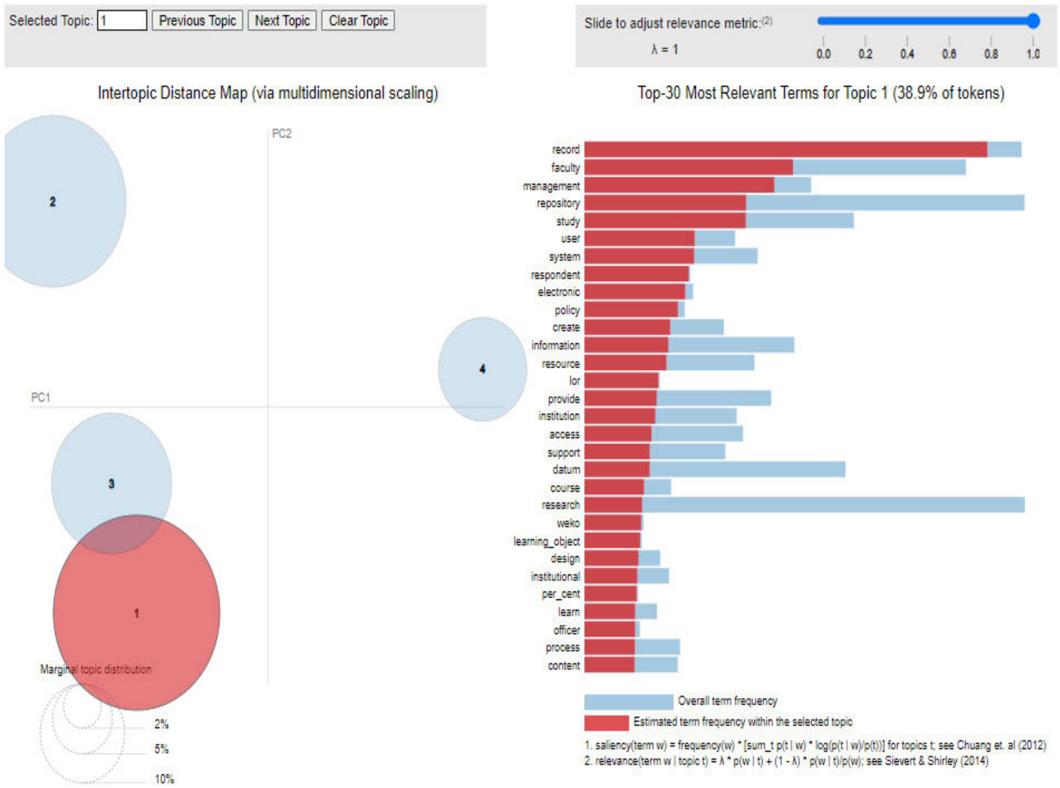

Figure 6.4. Visualization of inter-related topics

This meta-analysis shed light on the development, purposes, advantages, and adoption of research repositories by departments and faculties in HEIs. Faculty research repositories are gaining popularity for several reasons, including creating an open and reusable central repository of structured data collections, consistency, ease of use, straightforward handling and storage of research data, learning projects, scholarly outputs, and platforms for peer-to-peer collaboration in the same discipline.

6.3 Qualitative techniques used and definitions

The various analyses that were used to present the qualitative data results are defined in this section.

Word cloud

Word Clouds were used to demonstrate the most frequently used words. The larger the font implies, the more the term was used. This helps to identify critical areas/themes. By identifying tags, word cloud software analyzes word frequencies digitally and displays them using different font sizes and colours (De Paolo & Wilkinson 2014: 41). More frequent occurrences of a word, concept, or term will be displayed in the word cloud with a larger font size, while less frequent occurrences will be displayed with a smaller font size or not at all. To help readers understand relationships and meaning, the visual representation or graphic depicts patterns of words and phrases from the text (Cidell 2010: 514)

Figure 6.5. Word cloud of frequently used words in the analysis

Source: Field data (2022)

Tree Map

Tree maps were used to show the data (frequently used words) regarding the size of blocks. Hence, the larger blocks reflect those words mainly used. The entire map gives a holistic view of how data is placed regarding the reference size. A common technique for displaying hierarchical data is the tree map. Tree maps help display the overall hierarchy and the specific attribute values from individual data entries by recursively dividing the display area into rectangles based on the hierarchy structure and a user-selected data attribute (Tu & Shen 2007: 1287). Since the 1991 debut of the tree map, data from numerous applications have been visualized using tree maps (Johnson & Shneiderman 1991: 285).

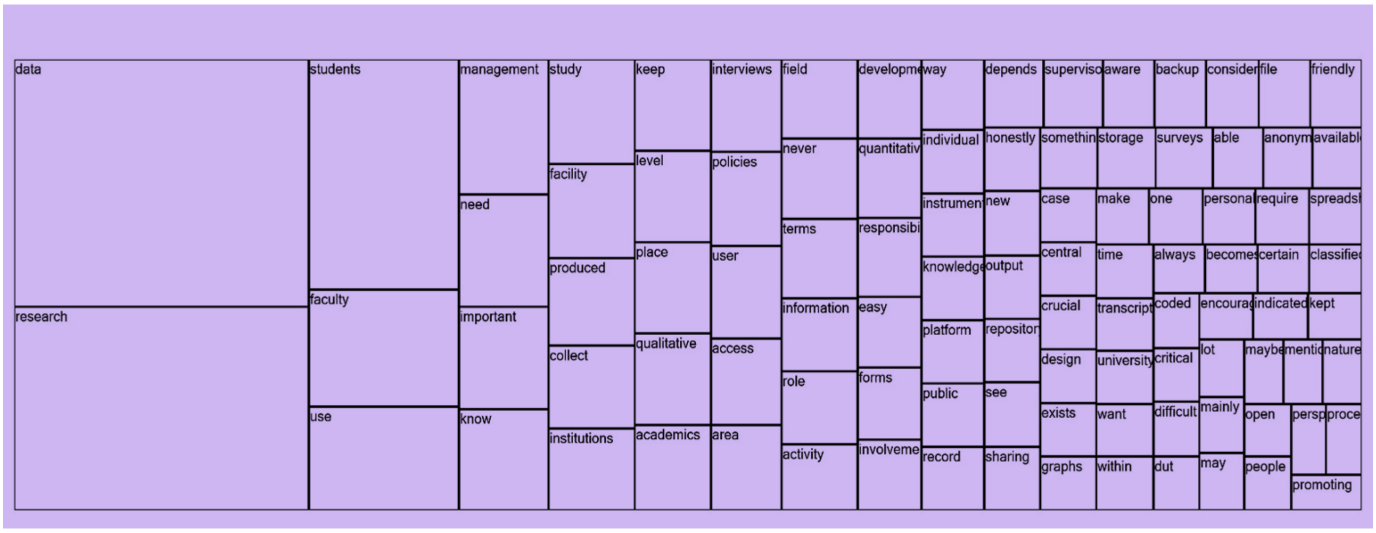

Figure 6.6. Tree map of frequently used words in the analysis

Source: Field data (2022)

Cluster Analysis

Cluster analysis was demonstrated using bubble diagrams. These diagrams illustrate the data (keywords) as 'bubbles.' The larger bubble size indicates the higher frequency of words/references. Furthermore, the closeness of the bubbles shows that there was a relationship between those words. One of the most beneficial steps in the data mining process for finding groups and spotting new, intriguing patterns in the underlying data is clustering (Frades & Matthiesen 2010: 81). Data objects are divided into groups (clusters) according to how similar or dissimilar they are using clustering algorithms. A valid cluster contains patterns that are more similar to one another than they are to patterns from other groups. Unsupervised, semi-supervised, or supervised methods can be used in clustering (Frades & Matthiesen 2010: 81).

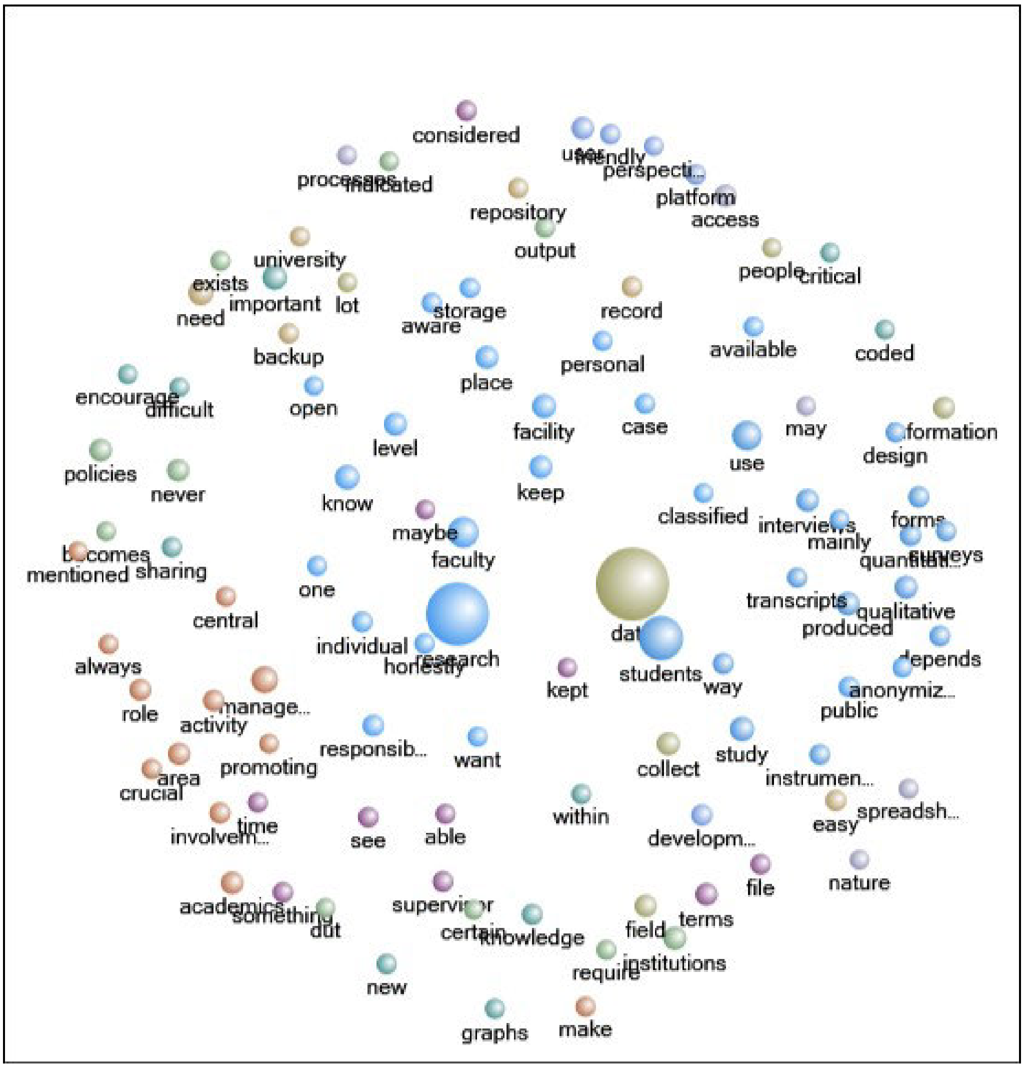

Figure 6.7. Clustering of frequently used words in the analysis

Source: Field data (2022)

Hierarchy Chart

Hierarchy charts were used to reflect the size of the nodes. The larger the size of the blocks implies, the more volume/concentration of responses in that area.

Figure 6.8. Hierarchy of frequently used words in the analysis

Source: Field data (2022)

6.4 Response rate

Overall, the expected number of participants from the interviews was 18, but only ten supervisors consented; therefore, the interviews were performed with the ten who agreed. Interviews had a response rate of 52.6%. ZOOM and Microsoft Teams were used to conduct all interviews online. To maintain their confidentiality, these participants' identities were not revealed throughout the analysis or presentation of interview data.

6.5 Themes of the study

The researcher gives context to the findings in this section by contextualizing them within the larger body of related literature on research data management that has already been discussed in Chapter Three and by tying the themes derived from the qualitative analysis to the theoretical frameworks underpinning the study. The Data Audit Framework (DAF) (Jones et al. 2009), Community Capability Model Framework (CCMF) (Lyon *et al*. 2012), and User-Centered Research Data Management Framework (UCRDMF) (Bugaje 2019) were the models that were adopted. In Chapter Two, these models are thoroughly discussed. The qualitative analysis generated four critical themes with many subthemes under each theme. The themes that emerged in this study are as follows:

- Theme 1: Importance of research in professional life
- Theme 2: Data Attributes
- Theme 3: Data Management
- Theme 4: Repository development

The study results are shown in the table below, grouped according to the themes that emerged after coding.

Table 6.2. Outcome of the coding process

CLUSTERING	THEMES	RESEARCH OBJECTIVE
Creating new knowledge, advancement, and progress, institutional mandates and policies, supervision, role and responsibility, trends, teaching and learning, HEIs, curriculum development, scientific objectivity, identifying gaps, minimizing mistakes, performance and recognition, community engagement, revenue	Importance of research	
Mixed methods, qualitative, quantitative, data formats, text, spreadsheets, graphs and tables, pictures, videos, file size, large,	Data attributes	Ascertain the management of the existing research data in a University of Technology Faculty.
Student responsibility, backup, location for storage, devices, access, processes,	Data management	Ascertain the management of the existing research data in a University of Technology Faculty.

The facility, research data classification, public, anonymous, sensitivity coding, risk levels, private, patents, commercial value, timeframes,	Research data access	Establish the usefulness of the faculty research data repositories
Research data reuse lack, platforms, systems, policies, research output, measurement, methods and research, reflections of research, more studies from the same data, support, and advocacy, sharing of research data within the faculty-university	Research data-sharing and reuse	Establish the usefulness of the faculty research data repositories
Central, duplication, training, and advocacy, free and funded, raw data storage, usability, and accessibility; storage and disposal; open source, multi-stakeholder, faculty-driven; guidelines, policies, verification, control, user-friendly, easy navigation, searching and retrieving, monitoring	development of repository structure for research data	Design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence

Table 6.2 shows how clustering groups emerged into themes. Themes are aligned with the research objectives.

6.6 Demographic profiling of participants

According to the demographic information provided for the interviews, all ten participants were from the Faculty of Accounting and Informatics. The distribution of participants across the various programs, including Business Information Management, Library and Information Studies, and Information Technology, was nearly equal. This suggests a variety of information from many programs. Interviews were conducted with both lecturers and senior lecturers. Most participants have only been employed at the faculty for five years or less.

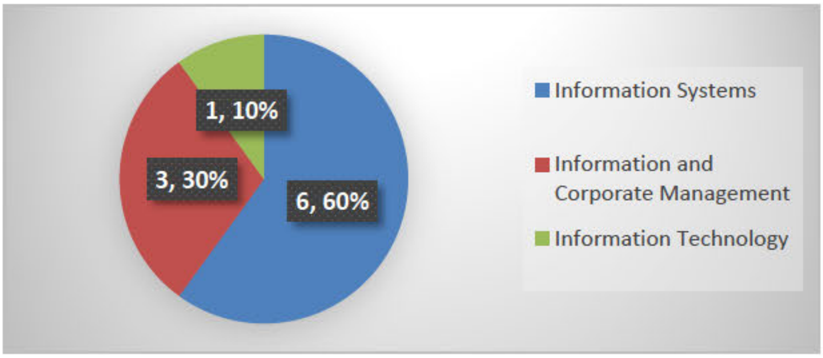

Figure 6.9. Participant demographic data

6.7 Importance of research

This section outlines why the research was seen as important by participants. Findings revealed included giving guidance on the research endeavors of their students within their respective departments. The researcher thought it was important to comprehend how the supervisors regarded the breadth and value of the research in their respective fields; every supervisor acknowledged the value of research in different ways. The findings revealed that participants elaborated on the importance of research in many ways. It promotes the creation and advancement of knowledge, addresses institutional mandates and policies that discourse research development, and supports curriculum development, teaching and learning, scientific objectivity, performance and recognition, financial gain, and community involvement, among other things. By discovering new knowledge, research advances knowledge and propels national economies. It also helps universities become more globally aware and provides them with a source of revenue. In their opinion, research aids academics in progressing in their careers through promotions to various higher university levels, such as Senior Lecturer, Associate Professor, and so on. The supervisors provided the following remarks regarding the importance of research:

> Participant 2: *"Research is important because we can create new knowledge through it. Without creating new knowledge, there will be repetition in what we do."*
>
> Participant 8: *"DUT sets the priority to research by enforcing it through policies, and as academics, we should comply with those policies."*
>
> Participant 6: *"Everything we do is confirmed by research; the role of teaching and learning has to do with critical thinking, which is the foundation of research. I believe I cannot entirely call myself an academic if research is not part of what I do and not just a part but a critical component of what I do."*
>
> Participant 5: *"Research counts because it brings revenue to the institution through the subsidies that result from research production in terms of outputs."*

Figure 6.10 presents the word cloud to demonstrate the most frequently used terms revealed as crucial when reflecting on the importance of research.

Figure 6.10. Importance of research

Source: Field data (2022)

6.8 Ascertain the management of the existing research data

In the DAF model, Jones *et al.* (2009) emphasize the significance of careful planning in how the data will be curated, including captured, concerning ascertaining management of existing research data. As a result, it is crucial to comprehend the formats in which research data is created and captured when examining research data management (Higgins 2011; Research Information Network 2008; Scott 2014: 121; Walters & Skinner 2011: 31). This is because data can vary significantly between research disciplines (Krier & Strasser 2014: 29; Ohaji 2016: 25). The interviews produced findings that provided the supervisors' view on managing existing research data emanating from their students' research projects. After coding, the results were grouped under the following themes: data attributes, data management, data location, storage, and backup.

6.8.1 Data attributes

In addition to using data attributes as its central focus, this study also looked at the existing research data forms, such as primary research data types generated by the guided students. The interviews revealed that the main types of research data varied according to the study undertaken. However, most revolved around mainly qualitative, mixed methods, and qualitative studies that participants supervised. Figure 6.11 shows a cluster indicating the higher frequency of words/references of data attributes revealed by the findings.

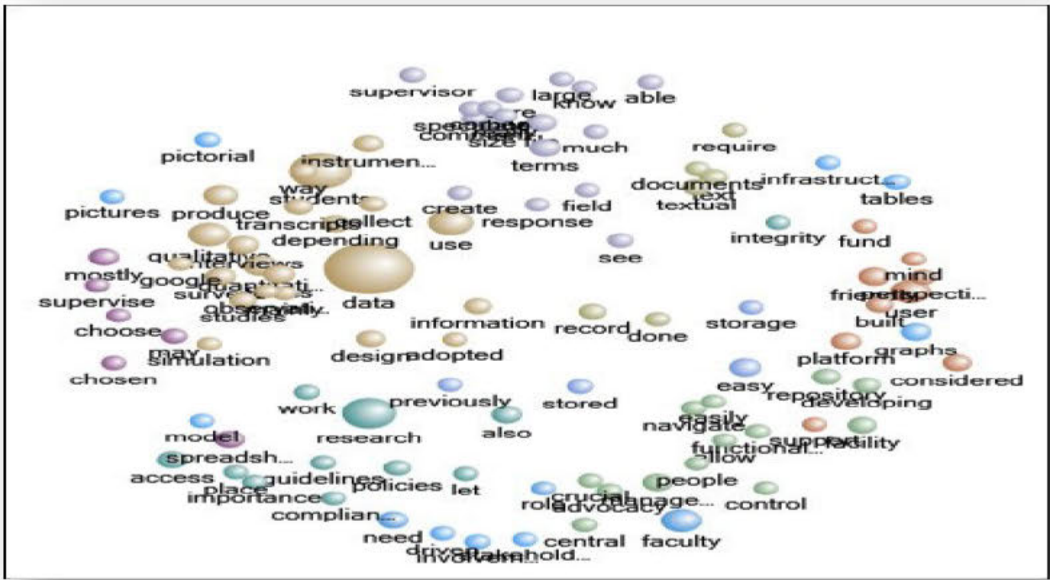

Figure 6.11. Data attributes cluster

Source: Field data (2022)

The results showed that mixed methods were the main emphasis of most participants. This demonstrated that academics were proficient in both research methodologies. Six participants indicated that their students' primary research types were based on mixed methods tools; therefore, the same study could be qualitative and quantitative. Some of the responses were as follows:

> Participant 1: *"Depending on the instruments that the students have adopted for their research studies, it has mainly been a bit of both, that is, qualitative data in the form of interviews, transcriptions, observations, and quantitative data in the form of surveys."*

> Participant 2: *"They produce qualitative and quantitative data depending on their studies. It may be a study where they use interview tools more on the qualitative side or send surveys through emails and quantitative Google forms."*

> Participant 6: *"My students collect qualitative and quantitative data, and they mostly work with designs and simulations of technological applications and systems, so as part of testing their products, they would use their participants in different ways, including interviews, focus groups, and surveys."*

Another four participants focused primarily on qualitative data derived via interviews. These were recorded and transcribed qualitatively. Below are a few quotes from the participants' interviews with the researcher.

> Participant 5: *"I would say for my students, and it's qualitative primarily because I am also a qualitative researcher, which influences the instruments my students choose. The area of research they choose forces them to produce qualitative data."*

> Participant 3: *"I am more into qualitative research, so mostly my students produce recordings from interviews they perform and later convert them into transcripts."*

> Participant 10: *"I am supervising two master's students who have produced qualitative data."*

The other two participants stated that they solely focused on quantitative research, so the primary data from their students was presented in quantitative formats. These were carried out via questionnaires and surveys. Here are a few verbatim comments from the interviews.

Participant 4: *"For my students, it's mostly quantitative research data by way of surveys using various platforms like Google Forms, QuestionPro, Survey Monkey, etc."*

Participant 9: *"I have had opportunities to supervise students who had produced quantitative data through the instruments they used."*

Primary research data formats

Research data formats are another characteristic that the analysis uncovered. The analysis put the primary research formats in chronological order based on participants. The hierarchy chart was used to analyze the research data formats primarily used by the supervisors' students. Figure 6.12 reflects the size of the nodes, which implies the volume/concentration of responses regarding the data formats identified.

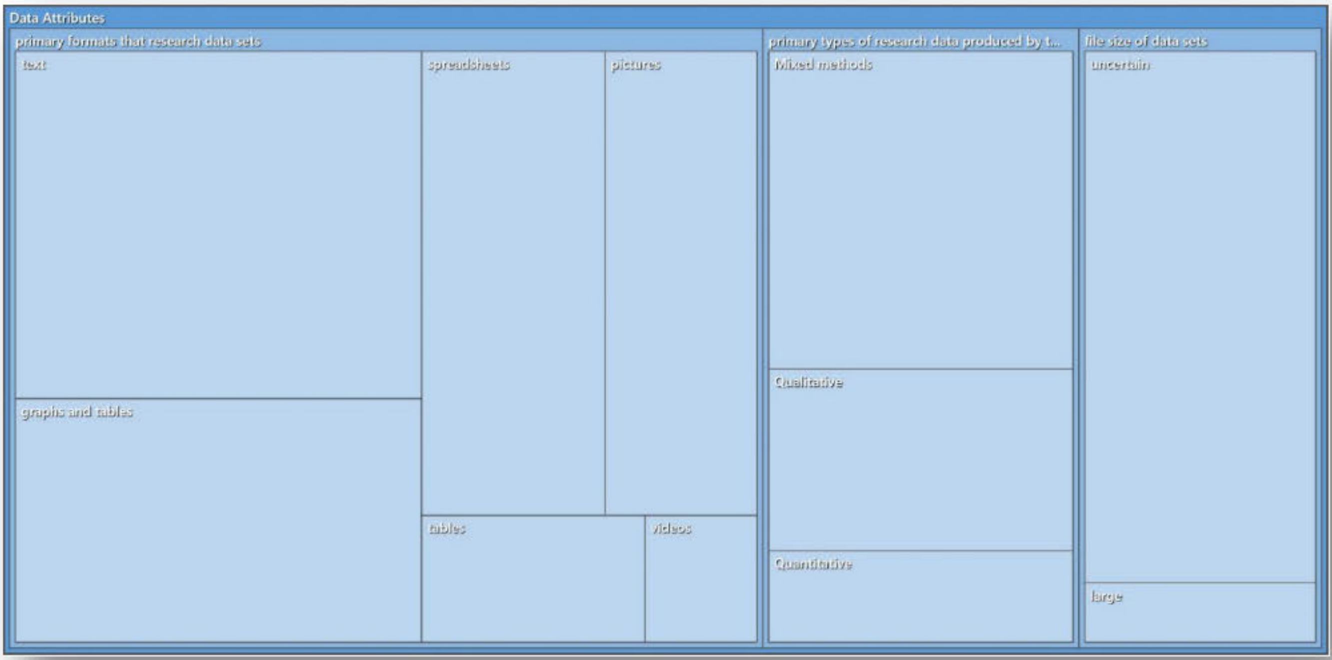

Figure 6.12. Data attributes hierarchy

Source: Field data (2022)

Text

The majority of participants indicated 'text' as being the primary format. This implies that most research data could be qualitative in text and document format. This is supported by the first participant, who said, *"It all depends on the instruments they are using, but most of my students' data is textual in the social sciences. Looking back at it, my students haven't done much data that will require pictorial information, videos, no, Yeah, so I would say the data primarily is textual".*

Spreadsheets

The findings revealed that most quantitative data (numbers) are kept on spreadsheets, making a quantitative perspective possible. As observed from the results, spreadsheets were ranked as highly utilized forms of research data. This is supported by the third participant, who said, "*spreadsheet files informed mainly by the nature of the research design a student may have chosen."*

Graphs and tables

The results ranked graphs and tables equally high as spreadsheets. This is logical, and spreadsheets allow for graphs and tables to be created from the data in their cells. Again, this reflects the high degree of quantitative data use.

Pictures and videos

As observed in the findings, some respondents indicated producing pictures as research data, but there were very few; therefore, they were not regarded as a highly ranked data format. The pictures as a research data format can be in various forms, such as images created from data, including diagrams, models, graphics, media, and schematics. The findings revealed that videos were the least type of research data format but were accounted for as they were mentioned in the findings as a research data format produced.

File size of research data sets

Most respondents were uncertain of how big the file sizes were. Nine out of the ten participants were uncertain and did not know. They asserted that the students would know better as they were custodians of their data. This shows the lack of knowledge of data attributes such as file size. The fifth interviewee explained that: "*I'm not too sure in terms of file size, and again it comes to the point that we will sit with the student and create the instrument, and the student goes to the field to collect data, but when that data comes up, we have not paid attention to file size really, but I would believe that some of these transcripts are not too large, it's a guessing job.*" Only one participant asserted that files were large but could not quantify byte size. The first participant's comment was as follows:

> "*I would not be sure, but as far as I know, I have not had a student that had asked for assistance may be due to producing extensive data that would require some type of hostage facility*."

According to the findings, most researchers produced their data in digital form, supporting the claims made by Cox and Pinfield (2014), Kahn *et al*. (2014) and Ohaji (2016) that the development of ICTs in academic and research institutions has fuelled the explosion in the production of research data in digital form. The results showed that fewer data was produced in videos and pictures. An ethical conundrum arises from using video recordings in human research because many subjects might not be open to having their responses captured on camera. As a result, video data in human research is not very common. The CCMF, according to Lyon *et al*. (2012), issues a caution that certain ethical obligations may place restrictions on the use of the data by researchers that go beyond the primary objective for which consent was sought and obtained from the participants.

6.8.2 Research data storage and backup

Since the primary focus of the study was the current management of research data, it makes sense that this was the most prevalent theme. Therefore, the sub-themes below inform it.

6.8.2.1 Current data management and backup process of the research datasets

The finding revealed a dire lack of data management and backup processes for research datasets.

a) Entirely student responsibility

The findings of the interviews revealed that currently, the data management process was entirely the student's responsibility. The respondents, as supervisors, did not get actively involved in the process. This showed a lack of academic/faculty involvement in research data management. The first interviewee indicated: "*I have no idea; since it's in the custody of the student and not the faculty or the university, I would not know. In this case, we are at the mercy of students because even as the supervisor after the students have graduated, we often also hold on to the thesis.*"

b) No proper backups

The findings of the interviews revealed that there were no proper backup procedures and systems. The results showed that participants had assumptions that students were backing up their data. The sixth interviewee commented, *"I would think there is some level of backup that students do, especially during the research journey, because they would need to reflect on the data now and then. After the submission of the research report, I cannot say if there is any backup and maintenance to the research data."*

c) Facilities and locations for storage of research data

The findings revealed a lack of facilities and locations for research data storage. All interview participants indicated that there were no storage facilities at the time the interviews were conducted, with one of the comments from the fourth participant presented below:

> *"Honestly, we do not have facilities to keep our research data in the faculty. Normally, students keep their research data."*

The participants further revealed that research data storage revolved primarily around students; therefore, it was the students' responsibility. The participants indicated that supervisors were only involved in research instrument development. This comment from the fifth interviewee reflects this: "*What supervisors mainly work on are the instruments before the student goes to the field to collect data. After the student has collected data, the results sit with the student, not with me as the supervisor or with some central office that makes sure these records are, or this data is kept safe and disposed of properly.*" The findings revealed that students only have access to research data because they store it on their own devices, which tend to include computers, phones, and external hard drives. Some students include their data in the written dissertation or thesis appendices. On paper, the student promises that the information will be stored securely, but there are no facilities to do so, which is not entirely the student's fault. The analysis also revealed a lack of proper processes to dispose of research data formally. *"There are no tangible processes to follow even when it comes to retention or destroying the data. It's up to the researcher to decide what to do with the data,"* stated the fourth interviewee. This is because everything about research data storage seems to be done individually and not centrally or formally. *"We keep it at the individual level; we also encourage our students to keep it for a certain period in case there is a need for going back to it,* " commented the fifth interviewee. The participants also shared that it becomes difficult to access research data after the research has been concluded because it is with the student. The fifth participant admitted that how each researcher stores their research data inhibits subsequent studies and publications. "*Without the data, it's challenging to convince students to publish, let alone supervisors*."

The CCMF and UCRDMF models state that technical infrastructure is required for data preservation, discovery, access, and collaboration. The results show a contradictory message, with the institution stressing the necessity for secure storage of research data while there is no implementation of the necessary management tools. The findings reveal that currently, there is no simple repository system for raw research data, only for completed theses, dissertations, and publications, among other research outputs.

d) Current classification process for students' research data

This sub-theme examined the current classification process, including codes and levels for student research data. The consensus among participants was that research data has always maintained anonymity. Another participant asserted that coding and classification depended on agreements and conditions between the institution and the respondents surveyed. Some of the classification levels were revealed to be request based, whereby data could be requested, and conditions and restrictions would be placed with relevant consent forms. In such cases, the fourth participant shared the processes involved in a statement: "*In this field, there is a level of restriction to information, but data can be made available upon request from the principal investigator. So, in such a case, only the metadata can be provided, not the actual data. The data is then made available upon request to ensure proper consent. So I would say, it depends on the nature of the study*." From a 'public' perspective, data were kept anonymous, and restrictions were placed for public use. The data were coded accordingly for public use, provided it was not sensitive. In the case of places such as hospitals, data would be needed to be coded for the sensitivity of medical information. The fifth participant shared the processes for sensitive data handling, "*I have studies where the focus is on hospitals, there is critical information, and anonymity is very important. In such instances, there would be a need for confidential coding of such data. Especially in small provinces where it would be easy to identify the institutions under study.* "Participants also shared that the critical aspect of risk levels had to be considered, which ultimately would inform coding and restrictions. There was a consensus regarding the 'private' data perspective; if the research leads to patent development, it must be kept private until registration and release of the patent. Similarly, if the research were to obtain commercial value, data would have to be kept private until the commercial value was explored and obtained.

6.9 Establish the usefulness of the faculty research data repositories.

Participants were asked to describe their faculty's steps to encourage researchers to utilize information created by other researchers or research institutions during the interview. The current data sharing and reuse perspectives and practices of this section were developed by data sharing and reuse, the mechanisms in place to ensure the digital preservation of research data, and the promotion of research data management practices in the faculty subthemes.

6.9.1 Current view of research data reuse

The findings revealed that the reusing of research data was found to be insufficient. The absence of systems and platforms to store research data was linked to the lack of reusing research data. Participants noted that this needed to be urgently addressed, with the first participant stating, "Maybe the argument is how we can share research data when we do not keep that *data. The research data that would have been produced is not available."* There was even a lack of policies to support or govern data reuse. The sixth participant felt that this area was neglected, saying, *"It's an important concept, but I would be lying if I said it's something that I have worked on or encouraged, not that I had any reasons not to, but it's an area that has been neglected.*" The seventh participant felt data reuse was mainly for research output, including publications and dissertations. However, the raw data that informed it was not reused, stating, *"It is something I have never considered; we have been more concerned with the research output, not the data that informs it."* The results also revealed that participants recommended more support and advocacy for data reuse. This was currently lacking in terms of awareness and the benefits associated with it. Reusing data can encourage new approaches to research and lead to discoveries in the following ways: it can offer a reliable roadmap of what has been accomplished thus far and what needs to be done; it can encourage measurement; it can support new methodologies; it can support more research. Instead of collecting data anew, more studies can be conducted using the same or comparable data; data reuse may also result in cost and resource savings for data gathering.

6.9.2 Promoting the sharing of research data within the faculty-university

The findings revealed that there was also very limited to no data-sharing. Current data-sharing practices were not strong enough. Some participants were clear that they have no active involvement in sharing research data. There was no direction or guidance from the faculty either. The first participant indicated that *"I will not say there is any specific activity that I engage in to encourage sharing of data; what I've tended to see is that there isn't a strong inclination by the faculty and maybe in the university to do research from the research data that already exists."* As mentioned in the data reuse theme, there were no facilities for data sharing. This made it challenging to encourage sharing. There is more focus on data collection than data storage and sharing. The area of sharing is hence severely overlooked. The first participant supported the above statement by saying, "*I think a lot of us are more inclined towards collecting primary data, meaning that I've got to go out to the field and get the data even in the instance where some of us get to use secondary data, that secondary data is not data that has come out of our research, I think it's an area we should look into because there's been an overemphasis on collecting data."*

Data-sharing can promote secondary research, which can be more convenient than only going to the field and collecting primary data. Other respondents believed that data management, including storage, currently seemed to sit outside of academia and had less involvement from academics. There was some limited sharing occurring perception based on feedback from two respondents who noted that; sharing data was logical if shared with people in similar research areas to allow for reproducibility. Otherwise, it would be of little to no value for those in different research areas; sensitive data should not be shared unless some access restrictions are implemented, and it should be anonymized.

6.9.3 Promotion of research data management and practice in faculty

This sub-theme examined the role of the respondents as academics in a faculty in promoting strong research data management and practice. There was a lack of promotion based on the following factors: most respondents indicated that promoting strong research data management and practice was currently absent. Some did not even hear of it from the faculty. Neither were they actively involved in it, with the first interviewee noting that *"There is not much role honestly, but as we are talking about this topic, I realize that there has been an assumption that students know about research data management and we as academics have not learned* *that we can also have a crucial role to play, after all as supervisors we are under scrutiny if the research is questionable.*" A severe lack of policies, procedures, and norms also drove data management and practice trajectory. Participants, as supervisors, never perceived themselves as participants in data management and practice. *"Honestly, I have not done much in that area; I have never realized that there can be a role that I can play as the supervisor. I think it's mostly because we have never seen ourselves as role players in research data management until now*," the sixth interviewee noted.

The CCMF and UCRDMF models stipulate that collaboration, data discovery, access, and preservation require technical infrastructure. Researchers can share data in a variety of ways, such as by "attaching data sets to published articles, depositing data sets in repositories, posting data on a personal or laboratory website, or responding to requests for data from other researchers" (Wallis *et al*. 2013: 2).

6.10 Design a novel digital prototype for a faculty research data repository platform based on research evidence.

This section looked at essential elements and functions that might guide the creation of a research data repository depending on the respondents' needs. A functional requirement declares a service the system must offer, support, or adhere to (Bugaje 2019: 105). Functional requirements relate to fundamental system operations and are gathered from current or potential system users (Maciaszek 2007: 75). Roles or connections to the system may be used to categorize users of the systems (Bugaje & Chowdhury, 2017). Researchers, funding organizations, and other primary, secondary, and tertiary users comprise the grouping. Depending on their position, they take on one or more of the following roles: a data consumer searches, browses, and downloads content from the repository; a data creator or holder uploads content sets access permissions, and provides metadata; a data administrator manages various database operations, such as, in some systems, checking user-uploaded datasets for errors to ensure that quality standards are met.

6.10.1 Repository Development

The following themes were identified and served as critical influences for repository structure development: location of the repository platform; usability and accessibility; stakeholder involvement; and functionality features.

Location

The participants' suggestions varied, with some proposing a central database for all institutional research data to avoid redundancy and duplication. In contrast, others suggested that each faculty have its sub-sections, creating a tiered structure. The first interviewee stated, "*I foresee one central data repository, but faculties have control over their sub-sectional (departmental) research data administration*." Meanwhile, the second interviewee stated: *I see this facility being developed in our faculty because the data we could have can help empower other students to avoid redundancy".*

Conflicting views about the location, with some respondents suggesting a central location while others wanted it at the faculty level for ease of use and control of research data. According to The University System of Georgia (2015), a faculty repository enables the sharing, access, and discoverability of the distinctive collections that faculties frequently have and are influenced by their programs. The second common reason for creating faculty research repositories is archiving and curation. It permits information to be reused and shared, reducing the cost and duplication of work in developing instructional materials. The dispersion of this data among different research institutes, departments, and individual researchers made accessing, sharing, and re-using it challenging, according to a study by Ng'eno (2018), even though Kenyan research institutes produced a sizable amount of research data. In the UK, where the current data infrastructure is dispersed across numerous faculties or sites, Brown *et al*. (2015) noted a comparable problem. According to Brown *et al*. (2015: 86), the disadvantage of this configuration is that it makes it challenging to create a central data management facility or unified storage solution. Implementing standalone data storage systems could lead to dissimilar or incompatible metadata standards, ultimately challenging interoperability between the systems.

Training and Advocacy

The findings revealed that participants believe there can be more centralized training and advocacy on the system, which can create awareness of the system across the institution. In support of this statement, the first participant commented, "*Training and advocacy should be done centrally, but faculties can load, manage, and use their research data guided by policies."* The fifth participant also emphasized the importance of training and advocacy: "*Advocacy and* *awareness would be crucial for people to understand how this platform would work,* as a comment.

There is a lack of training and advocacy because data sharing has only recently been introduced to the research environment. Matlatse (2016: 78) describes it as an emerging concept in South Africa. Tenopir *et al*. (2012: 71) reported similar findings, stating that 59% of American researchers claimed their universities did not provide them with guidelines for data management, which is why they lacked these skills. Regarding the CCMF, Lyon *et al*. (2012) emphasizes the importance of education and training for data management. Skills and knowledge are essential for carrying out the various tasks and actions in the data curation lifecycle.

Free and funded

The results revealed that participants believed the repository system was crucial and central to research data management. There were different views relating to the repository's location, whether central (to mitigate the financial responsibility of one faculty) or at the faculty level. The results also revealed the firm view that the ideal situation would be an open-source system that would not require funding. In cases where an open source was not an option, it was suggested that the university should pay the fees associated with the repository. To elaborate more on this, the third interviewee shared the following comment: "*It should be free; I don't think faculties would be willing to commit funds in these terrible times,"* while the eighth interviewee stated, "*Financial commitment is something the faculty should consider, and it would be great if the university can fund such an initiative."*

Building a reliable and robust data infrastructure has never been cheap. Since taxpayers fund public universities with limited and unviable funding, other infrastructure may be viewed as less critical because the limited resources are directed toward essential university operations. In a study investigating research data management practices in Malawi, Chawinga (2019) revealed that 62.7% of researchers cited a lack of data storage and network infrastructure. In a similar vein, Ng'eno (2018) observed that despite Kenyan research institutes producing a significant amount of research data, access, sharing, and reuse of this data was difficult due to its dispersion among various research institutes, departments, and individual researchers. Brown *et al*. (2015) noted a similar issue in the UK, where the current data infrastructure is dispersed across various faculties or sites. According to Brown *et al*. (2015: 86), the risk of such a setup is that it makes it challenging to develop a unified storage solution or central data management center. Implementing standalone data storage systems could result in disparate or incompatible metadata standards, ultimately making the systems' interoperability difficult.

Store raw data

The results revealed that participants believed that a central place was needed to keep raw data that can be accessible at a point of need. This can be used for other/future studies instead of collecting data again. However, restrictions and protocols will have to be put in place. "*We need a go-to central place for the faculty, where students can have access to store data during the research process, and the final raw data can be given certain access restrictions depending on its importance and confidentiality,"* the seventh participant stated.

According to Mi *et al.* (2021: 137), researchers have access to a repository to record and store data at various stages of a project, then maintain it over time and provide documentation of connections between projects due to the research repository's ability to offer raw content.

6.10.2 Usability and accessibility

The findings revealed usability and accessibility as crucial factors. The participants emphasized that for usability and accessibility to be achieved, the system must do what it intends to achieve. The infrastructure must be flexible to accommodate multiple data formats. "*Infrastructure that would accommodate all research data formats,*" as the ninth participant noted, cannot be rigid. Participants identified storage and disposal of research data to be the main aim and priority as that is the primary function of a repository; ethical guidelines prescribe that research data collected should be destroyed after a certain period to protect participants and respondents. To achieve that, the fourth participant commented, "the *university should also endorse that, and there should be a policy framework that will outline what should be stored or disposed of.*" The findings revealed no guidelines for all the above factors during the interviews. The participants recommended that systems be developed and define what can be stored and what can be disposed of. The tenth participant further suggested that open-source software platform design should be considered "*What needs to be considered is having such a facility through open-source means and the pro and cons of that."*

6.10.3 Involvement and stakeholders

The findings revealed that there should be high-level faculty stakeholder involvement. The development of such a system must be driven by faculty and not be a single student/staff responsibility. "*We need a faculty-driven model as faculties need to understand their role in research data management, and there should be compliance,"* the ninth participant noted. The ninth participant further indicated the need for multiple stakeholders' involvement from the faculty and the institution as such requirements and usability could be defined. The participants further recommended that the platform have a high level of integrity built into the system with appropriate policies and guidelines to guide the development trajectory. The seventh participant suggested that "*There should be guidelines on access, deposit, and use of the research data*" pertinent to access aspects, data types, regulations, rules and conditions of use. The participants further recommended that the system allow for data verification to ensure its validity, and access control would also be appropriate for a system such as this. Lyon *et al*. (2012) encourage research stakeholders to collaborate with other stakeholders in the CCMF because it improves the systems and processes for conducting research. Applying this logic to the current study, the collaborative effort between system developers and researchers in defining functionalities signifies a high-quality research repository platform.

6.10.4 Functionality and usability features for a suitable repository

A system's user interface is the collection of inputs and outputs that the user uses to activate its features (Satzinger *et al*. 2016: 219). Since the user interface is the primary communication between the user and the system, it is crucial to system functionality design. The following functionalities were seen as necessary for the system.

User-friendly perspective

Most participants commented on the importance of user-friendly system functionality; it was the most highly ranked factor and made a logical argument. These were some of the comments from the participants:

Participant 3: "*It should be user-friendly and built with the user perspective in mind*."

Participant 4: "*It should be a user-friendly built platform*."

Participant 5: "*User perspective should be considered when developing such a repository*."

Participant: "*User-friendly interface*."

According to Bugaje (2019: 105), not only must the various user groups be identified to achieve maximum system usability, but it is also crucial to fully comprehend each group's unique traits and the tasks each performs on the system. A research study by Taylor *et al.* (2015: 244) revealed that users' information needs and, implicitly, their data needs tend to be ambiguous, poorly expressed, and frequently only apparent at first glance. Additionally, system developers should be aware that the user's level of system knowledge can range from extremely basic to extremely advanced (Taylor *et al.* 2015: 258). Therefore, systems must be created to enable even the least technologically savvy user to search for and find data efficiently. Many systems fail due to a lack of user-friendliness. Hence, the system should be easy to use and user-friendly, and this can be attained by developing the system according to the users' perspective.

Easy navigation

The findings revealed that participants regarded navigation as a critical factor in system design. The system should allow for easy navigation. This is important, especially since it would store numerous large datasets, and users will want to navigate through them quickly. Some comments are indicated below:

> Participant 1: *"For these repositories to be adopted, their functionalities must allow people to navigate them to find information easily."*
>
> Participant 2: *"I would support the development of this facility within a faculty not only to have records of research data but also for ease of reach in terms of faculties as content owners."*

A study by Bugaje (2019: 144) revealed that users preferred not to open new tabs or leave the current page to quickly learn key details about search results as they scroll down.

Re-usability

The ultimate objective of data management, sharing, and preservation is re-usability. Metadata has been observed throughout the work as the necessary information for data repurposing or reuse. The findings revealed that participants recommended that the system allow data to be reused, not just for storage. Hence this could include raw data and statistics that can inform other studies. These are some of the statements by participants:

Participant 2: "*It will be easy to refer students to the data for reuse.*"

Participant 6: "*User perspective should be considered when developing such a repository as it will be for the user to use, not just a storage facility. I would have a user in my mind when I do it.*"

Kabanda *et al.* (2023: 10) revealed three excellent reasons for adopting a reusability approach to research data management; they especially deserve consideration due to the dynamic nature of technology and information systems and the requirement to maintain the usability and accessibility of stored data. Throughout their entire life cycle, digital materials are fragile and susceptible to change brought about by technological advancements. The actions taken—or not taken—at each life cycle stage directly impact researchers' capacity to manage and preserve digital materials at later stages. Digital materials cannot be reliably reused unless carefully curated to preserve their integrity and authenticity.

Searching and retrieving

In this section, the participants commented on the importance of searching and retrieving research data effortlessly. As observed in the participant's comment below, this is primarily due to students wanting to search for specific material amidst high volumes of existing data.

Participant 9: *"simple searching and retrieval functionality."*

Witt (2008: 198) claims that research data discovery systems outline points of access required for searching and perusing data repositories and strategies for assisting users and user agents in finding data.

Access control and monitoring

Accessibility includes details about data access rights, rules, and guidelines that inform users of access points. The system uses the same metadata to grant access to specific users or machine clients. When choosing a repository for a faculty to enable search, access, and discoverability of content, these critical factors are essential to consider (Shakeri 2013: 79). The findings revealed that participants recognized the importance of access control as a crucial factor for security, tracking, and integrity, and monitoring the use of the platform can inform measurement and improvements.

Participant 4: *"The system must provide functionalities with well-defined access and control."*

Participant 5: *"We should be able to monitor the use of this platform."*

Users of research data management systems interact with systems for various purposes, with varying levels of engagement, skill, and ability, as do users of systems generally and from diverse personal and disciplinary backgrounds (Bugaje 2019: 119). These peculiarly combined variables for each user impact the overall user experience and system satisfaction.

6.11 Appraisal of the Chapter

This section presents the study's significant findings, as shown below in Table 6.3, and closes off with an appraisal.

Table 6.3. Summary of study findings

The study findings revealed	**Findings**	**Objective**
Importance of research	It supports the development of curricula, teaching and learning, scientific objectivity, performance and recognition, financial gain, and community involvement while advancing the creation and advancement of knowledge. It also addresses institutional mandates and policies that discourse research development.	
Data attributes	The findings indicated that participants worked with mixed methods research data more frequently.	Ascertain the management of the existing research data

	Most participants indicated that text was the primary format, followed by spreadsheets and graphs, with pictures and videos being the least frequently used. The majority of respondents weren't sure how large the file sizes were.	in a University of Technology Faculty.
Data management	There were no proper backup procedures, systems, facilities, or locations for storing research data, so the data management process fell entirely on the students. Participants' anonymity with research data was always maintained during the classification process.	Ascertain the management of the existing research data in a University of Technology Faculty.
Research data access	There was insufficient use of research data. The lack of systems and platforms for storing research data was associated with a lack of data discovery.	Establish the usefulness of the faculty research data repositories.
Research data sharing and reuse	Additionally, there was very little to no data sharing, and some participants explicitly stated they were not actively involved in sharing and re-using research data.	Establish the usefulness of the faculty research data repositories.
Development of repository structure for research data	A central location was suggested, but some participants preferred it to be at the faculty level for ease of use and control of research data. Centralized training and advocacy on the system could raise awareness throughout the institution, and an open-source system wouldn't require funding. When an open source wasn't an option, the	Design a novel digital prototype based on research evidence for a University of Technology faculty research data repository platform.

	university should cover the costs of the repository; High-level faculty stakeholders involved, user-friendly system functionality that is simple to navigate, flexible infrastructure to accommodate various data formats, access control as a critical component for security, tracking, and integrity, and monitoring platform use	

In chapter six, the researcher conducted a comprehensive meta-analysis to explore the motivations and factors driving the establishment of Higher Education Institution (HEI) research repositories. The analysis focused on crucial research papers related to research repositories, providing a detailed understanding of their creation and function. Additionally, the chapter delved into a qualitative analysis of research supervisors within the Faculty of Accounting and Informatics, focusing on key study themes such as research data management, the value of faculty research data repositories, and the development of a modern digital prototype for a research data repository platform based on research evidence.

The chapter highlighted how supervisors perceive the value of research and its management, presenting findings aligned with the study's research questions and themes. Notably, it emphasized the prevailing patterns of research data management within the Faculty of Accounting and Informatics, emphasizing the lack of centralized systems for managing research data. The study identified that individual researchers primarily handled research data, often stored on personal laptops, emails, or external hard drives. The absence of investment in research data management at the faculty level contributed to these challenges.

Participants acknowledged the benefits of effective research data management, including scientific advancements, cost-effective data reuse, and enhanced accessibility for data reusers. However, the chapter revealed that researchers were not actively sharing their data or reusing data from other researchers due to the absence of centralized platforms facilitating such practices. Despite a desire to share their research findings, the lack of a structured system hindered effective data sharing and reuse among researchers.

In summary, this chapter provided valuable insights into the current state of research data management within the Faculty of Accounting and Informatics, shedding light on the prevailing challenges and the need for investment in centralized platforms to optimize data sharing and reuse, ultimately advancing scientific progress.

CHAPTER SEVEN: SUMMARY, RECOMMENDATIONS, CONCLUSIONS AND IMPLICATIONS OF THE STUDY

7.1 Introduction

The study aimed to investigate research data management in a faculty context to support researchers by designing a faculty research data repository platform conducive to a University of Technology. The first section of this chapter provides a recap of the study's six chapters. The synopsis of the results and recommendations that contributed to the study results are further described in the chapter. This chapter also evaluates the success of the study's intended objectives and the potential effects of its findings on society. The significant contributions are then articulated, and proposals for further research are presented. The chapter summary and conclusions are represented graphically in the explanatory sequential mixed method study diagram the researcher utilized, as illustrated in Figure 7.1.

Explanatory Sequential Mixed Method Study

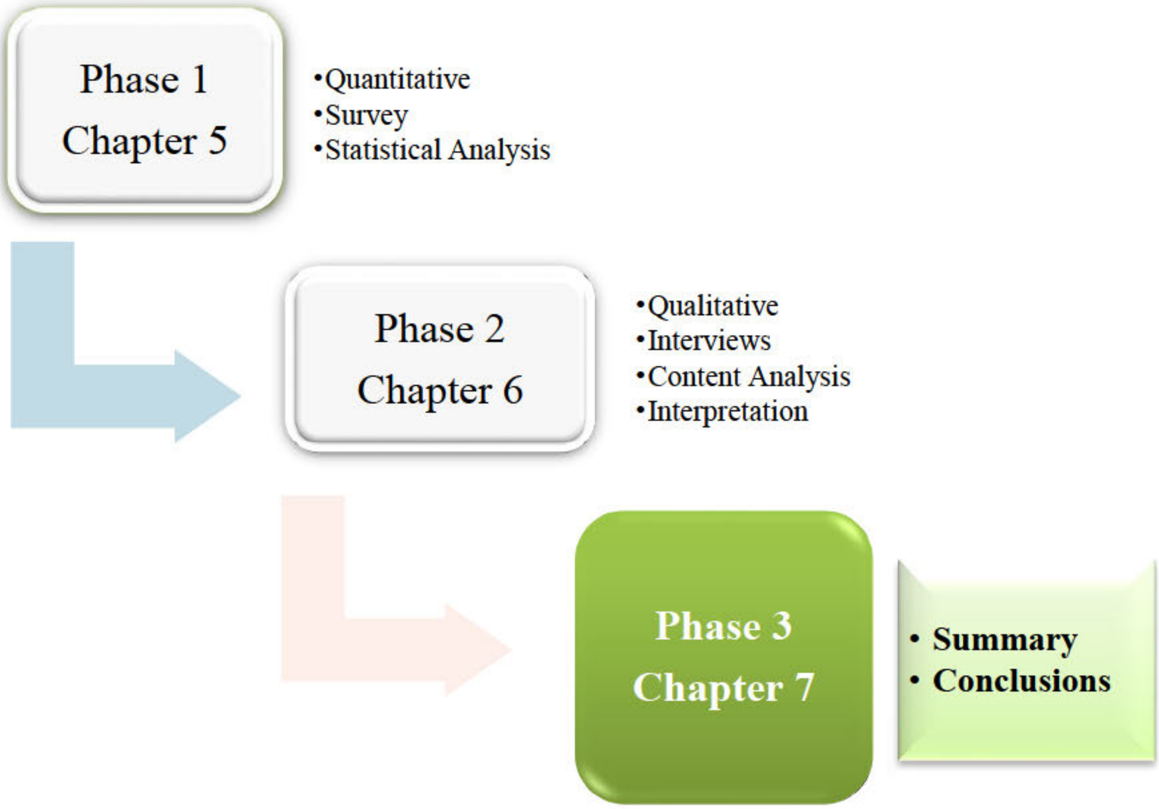

Figure 7.1. An adaptation of Creswell's Explanatory Sequential Mixed Method Study

Source: Creswell and Plano Clark (2018)

7.2 Summary of Study

This section summarizes the research, outlining its goals and objectives and the models that served as the study's lens and literature review. It also showcases the data collection methods, techniques for analyzing data, and data interpretation methods that helped shape the study. The study summary is captured in Table 7.1, providing the entire study's alignment.

Table 7.1. Alignment of the Research

TITLE: The Design of a Faculty Research Data Repository Platform conducive to a University of Technology			
AIM: The aim of the study is to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology			
Research Objectives	**Theoretical Framework**	**Data Collection Methods/Data Sources**	**Data Analysis/Results**
[RO 1]: *Analyze the most pertinent studies on research repositories at HEIs.*		Literature Review ***Chapter Three***	Published research output: *A systematic review of faculty research repositories at higher education institutions*. Latent Dirichlet Analysis ***Chapters One and Six***
[RO 2]: *Ascertain the management practices of the existing research data in a University of Technology Faculty*.	Data Audit Framework	An online questionnaire with registered Masters' and Ph.D. students for Accounting &Informatics Faculty Online interviews with Masters and Ph.D. Supervisors in the Faculty of Accounting and Informatics	Statistical analysis using SPSS Using NVIVO for qualitative data analysis Latent Dirichlet Analysis Faculty Research Data Repository Platform Model ***Chapters Five, Six, and Seven***

		Chapters Four, Five, and Six Annexures H &I	
[RO 3]: *Establish the usefulness of the faculty research data repositories.*	Data Audit Framework Community Capability Model Framework	An online questionnaire with registered Masters' and Ph.D. students for Accounting &Informatics Faculty Online interviews with Masters and Ph.D. Supervisors in the Faculty of Accounting and Informatics ***Chapters Four, Five, and Six*** ***Annexures H &I***	Statistical analysis using SPSS Using NVIVO for qualitative data analysis Latent Dirichlet Analysis Faculty Research Data Repository Platform Model ***Chapters Five, Six, and Seven***
[RO 4]: *Design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence*	User-Centered Research Data Management Framework	Literature Review An online questionnaire with registered Masters' and Ph.D. students for Accounting &Informatics Faculty Online interviews with Masters' and Ph.D. Supervisors in the Faculty of Accounting and Informatics ***Chapters Three, Four, Five, Six, and Seven*** ***Annexures H &I***	Statistical analysis using SPSS Using NVIVO for qualitative data analysis Latent Dirichlet Analysis Delphi methods Faculty Research Data Repository Platform Model ***Chapters Five, Six, and Seven***
Integration of Research Project Summary, Conclusions, Significant Contributions, Implications, and Future Research ***Chapter Seven***			

This study examined research data management in a faculty context to support researchers by designing a data repository platform compatible with the university's culture (University of Technology). After setting the aim and objectives for the study, a robust theoretical framework was sought. Three models, namely the Data Audit Framework (DAF), the Community Capability Model Framework (CCMF), and the User-Centered Research Data Management Framework (UCRDMF), served as the study's guiding principles. Each model offered detailed instructions, such as that DAF was the best model for determining research data management. CCMF detailed the roles, responsibilities, and requirements for RDM's effectiveness and efficiency as a crucial component of faculty-generated research data. UCRDMF demonstrated the development of a simple prototype for a user-focused, data-conscious research data management system.

The literature review was organized around the study's goals and focused on using the review synthesis method to analyze the most relevant studies on the research repositories at HEIs. The literature review also covered broad topics like research data creation, evaluation, preservation, disposal, and storage facilities to determine how to manage the existing research for researchers. To address critical issues like research data sharing and reuse by researchers, a literature review was conducted to assess the usefulness of faculty research data repositories. The research objective, to design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence, was informed by examining the existing research data repositories, design and development, metadata standards, and interoperability. The literature also summarized the RDM challenges faced by the research communities.

Pragmatism served as the philosophical foundation for the study, and the mixed methods research method was used as its investigative strategy. To gather quantitative data for this study, open-ended and closed questions from an online questionnaire and interviews were used for qualitative data. Quantitative data were collected through a Questionpro survey, and descriptive statistics in percentages and frequencies were carried out using SPSS version 28. Thematic analysis was carried out using the NViVO software on qualitative data gathered using Zoom and Microsoft Teams, which were used for conducting the interviews. The study also included content analysis to support and supplement data from the questionnaire and interview results and to help understand the phenomenon of interest more thoroughly.

The presentation and discussions of the data collection technique used for this investigation were also provided, highlighting that research data management was an individual's responsibility. For example, 91.2% of researchers tended to save their data on personal, email accounts, and office computers. Faculty had not engaged in research data management, and 71% of researchers did not share their data since there was nowhere to put it. This was the primary driver of the issues. The presentation and discussions of the qualitative method focused on the key themes of the research, which were informed by the research objectives, and they revealed patterns comparable to those of the quantitative methods. Researchers were prevented from sharing their data or using other researchers' research data, primarily because no centralized platforms allowed for these activities.

7.3 Conclusions of Study

The results for the four research objectives covered in this study are presented in this section systematically.

7.3.1 [RO 1]: To analyze the most pertinent studies on research repositories at HEIs

In line with the PRISMA method, the first research objective gave an in-depth analysis of pertinent studies on research repositories at HEIs. Eleven articles that covered the review's scope were subjected to quality analysis. The objective of the evaluation was to learn more about the advantages of faculty research repositories, their functions, and their usage. The analysis of the corpus of eleven articles, using the clustering algorithm of Latent Dirichlet Analysis (LDA), revealed significant vital phrases about the research data management ecosystem, such as “research repository," "research publication data," "participant data," and "record management." This meta-analysis clarified HEI departments and faculties' origins, functions, benefits, and uptake of research repositories. The analysis revealed that faculty research repositories were becoming more common for various factors. Research repositories were found to be suitable for departments or faculties who intended to build accessible and reusable centralized databases for systematic collection of data; continuity; the convenience of use; precise storage, and management of research data. The analysis also revealed that the repositories were suitable for capturing educational projects and intellectual outputs and could be used as platforms for peer-to-peer collaboration in the same field. Based on the analysis, it is recommended that a faculty research repository be established so that lecturers and students can take full advantage of the benefits of having such an infrastructure on-site

7.3.2 [RO 2]: To ascertain the management practices of the existing research data in a University of Technology Faculty

The importance of determining the management of research data, including how it will be aggregated and recorded, is emphasized by Jones *et al.* (2009). Therefore, when examining research data management, it is essential to understand the formats in which the data is created and captured. The results of the research objective were presented based on the themes unveiled by the research instruments. For instance, quantitative and qualitative findings showed that digital texts were the most produced of all data attributes (Figure 5.5, Chapter 5 and Figure 6.12, Chapter 6). Most respondents generated observational data, so digital text comprised most of the data. The data management and backup theme revealed that survey respondents stored their research data on personal, email, and work computers and that their supervisors had no control over the data produced by students because it was entirely in their hands.

Additionally, the study found that most respondents' data (59%) was less than 1 GB, likely due to the types of files produced. The results showed that most student-generated data fitted into the least protected classification category. This category required little to no categorization; meanwhile, the supervisors mentioned that research data was always kept confidential. A repository for the faculty to necessitate proper management and backup of research data was recommended as a platform that would assist the faculty to be aware of the content at their disposal.

7.3.3 [RO 3]: To establish the usefulness of the faculty research data repositories

One of research data management's objectives is the capacity to facilitate data exchange. The findings of this study divulged data-sharing behavior patterns, intrinsic motivators, technological advancements, and variables impacting data-sharing. According to the survey findings, 57.4% of respondents actively participated in research data-sharing operations. Additionally, 32% of respondents shared data because it was ethically sound to do so and because doing so sped up knowledge expansion. At the same time, 27.9 respondents declared to share data because open-access advocates advised them to do so. Contrarily, the results of the interviews showed that data-sharing procedures lacked rigor and that supervisors were not actively involved in disseminating research data. It was discovered that respondents used various tools to share data, including flash drives, emails, journal websites, and cloud storage (such as Google Drive and DropBox). The study did show that some barriers to extensive data sharing existed, though. Aside from the personal storage each researcher used, researchers had no other place to store their data, which limited extensive data sharing. The findings also demonstrated the lack of protocols or data-sharing standards, making it challenging for researchers to share their data. The lack of metadata prevented the data from being found, making it difficult to access.

The willingness of researchers to deposit their research data in a central repository was contingent upon their data being cited and used by other disciplines. The following elements were identified as inhibitors to research data reuse: complexity in discovering, locating, or gaining access to reusable data; legal or moral restrictions; insufficient data descriptions or metadata. Regarding research data preservation, 42.6% of respondents indicated that storing research data for three to five years would be beneficial and evaluating it at that point because it would still be valid. The recommendations included establishing an infrastructure for research data storage and preservation, which will allow discovery and archiving. The development of a research data management framework that will provide guidelines and procedures that would outline the processes involved in managing data was also recommended.

7.3.4 [RO 4]: To design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence

Through this objective, the needs of the respondents, who served as the creators of the repository for research data, were revealed, as well as the fundamental components and features that drove its development. The repository's location, usability and accessibility, stakeholder involvement, and functionality attributes were outlined as themes that significantly impacted how the repository framework was developed.

Regarding the platform location, the findings revealed that all students who participated in the survey agreed that a faculty repository was necessary. The findings revealed varied opinions from supervisors. Some believed it should be centrally located to eliminate issues resulting from funding commitments. In contrast, others insisted it should be premised in the faculty to enable the ingestion, access, and discovery of research data aligned with the faculty subject fields of study.

The findings showed a lack of training and advocacy because data sharing was primarily a solitary individual activity, not established at the faculty level, and was only recently introduced to the research environment. The results showed that respondents would favour a centralized training and advocacy strategy that could spread awareness of the system throughout the institution. This would also be helpful if faculty repositories were emulated throughout the faculties of the institutions. The findings were also in favour of starting an unfunded open-source repository. It was proposed that the university should cover the costs of the repository if an open source wasn't an option.

The results showed that the respondents needed a platform to store raw data and make it available when required so that it could be utilized in subsequent studies rather than having to collect it repeatedly. The findings identified usability and accessibility as critical aspects. The respondents desired a platform that could handle various research data formats, with the preservation and disposal of research data as a repository's primary function. Additionally, respondents wanted ethical guidelines to specify that research data obtained should be destroyed after a specific time while maintaining confidentiality during the destruction process.

According to the research findings, creating a faculty repository should involve high-level faculty stakeholders rather than solely the responsibility of lecturers, students, and staff. Since the user interface serves as the primary means of communication between a user and a system, respondents suggested the following functionalities as being essential: -

- The system must be simple and constructed from the user's perspective.
- Users want quick results, so the system needs to be simple to use since it will likely hold many different formats of data sets.
- The system should be designed for ingestion and discovery to allow research data reuse.
- Finding data should be simple and quick.
- Access control should be prioritized in the system and recognized as essential for security, tracking, integrity, and monitoring.

7.4 Recommendations of the Study

Based on the outcomes and suggestions, the following repository structure is recommended. To help with the design of the prototype, the recommendations recognized the system's key players. The following participants were noted as pivotal role players:

- The person who created or holds the data, whose primary tasks involve uploading data, configuring access privileges, and providing metadata;
- The information user, whose primary tasks involve looking through and retrieving data from the repository;
- The database's back end is under the control of the data administrator, who may also check client datasets for flaws in some systems to guarantee that quality requirements are fulfilled.

The study's results provided the respondents with crucial elements that are expected to form part of the main functions or features of the envisaged faculty research data repository platform. The researcher depicted the required system functions in Table 7.2 and described the role players and their expected actions in the platform.

Table 7.2. System requirements end-user recommendations

Sytem's Functions	Data Creator	Data Consumer	Data Administrator	Findings recommendations	Prototype execution
Ingestion	Author(s), Synopsis, Search terms		Quality checks	flexible to accommodate multiple data formats and sizes	Data Uploading
Searching	Can search using Search terms, author name, digital object identifier	Uses Search terms, author name, and digital object identifier to find content	Performs Quality checks; basic searching to test the system delivery	Easy to search	Searching
Browsing	Uses browsing To view content available in a department /by an author/ or by faculty	Uses browsing To view content available in a department /by an author/ or by faculty	Performs Quality checks for redundancy, etc	Easy navigation	Searching
Sorting		The user uses this function after receiving results from the initial search		Easy to use	System built-in

Metadata	To ingest	To use it to discover data/ to understand the data description	Compliance measures Ensure that metadata is attached to data records	Easy to discover	System to display metadata field by default based on the standards it has adopted
Usage tracking	Ability to check the usage of their deposited data	An indication of views and downloads for the consumer to have an idea of the usage/ popularity	An administrator can generate usage		System built-in
Data Security	Ensured by the owner by granting appropriate access restrictions to the data	Adherence to data restrictions	Monitoring	Security and integrity of the system	It is the data owner's responsibility; administrators should monitor and track; the system should allow monitoring functionality.
Compliance with service norms and regulations	Adhere to regulations		Ensure compliance	Compliance with standards and guidelines	The system should follow norms and regulations based on the standards

Source: Researcher's construction

7.4.1.1 The Delphi Methods

The researcher considered this mixed-method study's analysis, findings, and discussions. A clear conclusion from this study was the requirement to create the first platform for a faculty research data repository in UoTs and the entire South African university system. The researcher consulted LIS specialists on research data repositories using the Delphi method informed by Gallotta *et al.* (2018: 231-241) to create a conceptual model for designing a faculty research data repository platform for a UoT context in South Africa. LIS specialists are those with extensive knowledge of the design architecture of research data repositories, long-term preservation, and the concepts of open access and controlled access. Examples include seasoned managers, administrators, and librarians of research data repositories. To meet the faculty research data management needs of the UoT under study, it was essential to take advantage of their expertise when designing an efficient research data platform model. Throughout the Delphi procedure for this study, the researcher served as a facilitator. Based on anonymity, participants who were highly experienced experts in the LIS field were chosen. This ensured that participants' viewpoints during each phase of the Delphi process remained impartial or independent. Three gathering rounds were in succession.

The **first round** consisted of an open-ended questionnaire (Annexure L) to generate concepts for the research data platform model design. The facilitator created an overview report after participants finished and returned their responses. The experts from the LIS field who took part in the Delphi process received the summary report. The facilitator ensured participant identities were not revealed in the summary report.

A brief survey (Annexure L), used in the **second round**, was designed as a 5-point Likert scale with statements. The questionnaire was created after analyzing the initial round's data. The facilitator edited the first round's answers for being repetitive and inconsequential. Statements that shared content from the initial round were also included. The responses from the second round indicated that there was beginning to be some degree of agreement on the model for the faculty research data repository. Confidentiality was maintained by the facilitator when sending a summary report to the experts.

The **third round's** questionnaire was created with concepts from the second round in mind. Consequently, the third round's questionnaire was similar to the second round's (Annexure L). A decision was made at this point regarding the efficient design of a faculty research data repository for a UoT context in South Africa by experts in the LIS field. The projections demonstrate that experts generally agreed on the model's design. The design of a faculty research data repository for a UoT context was influenced by the Delphi method, data analyzed from this mixed method study, and the review synthesis of HEIs research repositories.

7.4.1.2 Proposed model for the Faculty Research Data Repository Platform

The conceptual framework provided here offers detailed instructions regarding how to plan and execute the design for the Faculty Research Data Repository Platform. The model will let faculties look into their users' needs, the current data, the place it's kept, and precisely who is in charge. Based on their findings and suggestions, they can lay out a platform for handling and sharing this data in a controlled setting with other faculties and institutions. This conceptual framework is founded on scientific exploration of research data management practices in the faculty under study, a literature review, proposals derived from the results, and expert advice derived from the Delphi methods. The model embraced some elements from the models that served as the director for this study both directly and indirectly because the evolution of current knowledge inspires the development of new knowledge.

Table 7.3 presents four characteristics of the proposed conceptual model with either present or absent components compared to the models that served as the study's guiding framework.

Table 7.3. Proposed components of the model compared with selected models

Recommended 1st Component: Information Analysis						
Sub-components			**Specifications in comparison to the models used in this model**			
			DAF		**UCRDMF**	
Sub-component	**Elements**	**Observations**	✓/✗	**Observations**	✓/✗	**Observations**
Meta-analysis of research repositories in HEIs.	Latent Dirichlet Analysis	The analysis revealed that faculty research repositories were becoming more common for various factors; a sound recommendation to establish one was made.	✓	Examination of departmental data collections and the procedures they follow regarding data curation and preservation.	✗	Emphasizes recognizing and defining the context of repository systems at large. It does not focus on research data repositories for faculties.
Quantitative	Web Surveys	Quantitative analysis revealed a lack of a central platform to manage research data for the faculty.	✗		✗	
Qualitative	Online Interviews	Qualitative analysis	✗		✗	

		revealed a lack of a central platform to manage research data for the faculty.				
Recommended 2nd Component: Design Criteria						
Sub-components			**User-interface-informed system capabilities**			
			DAF		**UCRDMF**	
Sub-component	**Elements**	**Observations**	✓/✗	**Observations**	✓/✗	**Observations**
Flexibility		The user experience is the primary interaction between a user and a system; these functionalities were identified as critical.	✗	The model does not either fully or partially cover these attributes	✗	The model does not either fully or partially cover these attributes.
Easy search functionality			✗		✗	
Easy navigation			✗		✗	
Security and system integrity			✗		✗	
Compliance with standards and guidelines			✗		✗	
Recommended 3rd Component: Faculty Research Data Repository						
Sub-components			**Location and Preservation processes**			
			DAF		**CCMF**	
Sub-component	**Elements**	**Observations**	✓/✗	**Observations**	✓/✗	**Observations**
Location		The study revealed a lack of a central platform to manage research data for the faculty	✓	Examination of departmental data collections and the procedures they follow regarding data curation and preservation.	✗	This element is not implicitly or explicitly covered by the model.
Retention period			✓		✗	
Recommended 4th Component: Faculty Research Data Repository Design						
Sub-components			**Repository Functions**			
			CCMF		**UCRDMF**	

Sub-component	Elements	Observations	✓/ ✗	Observations	✓/ ✗	Observations
System Functions	Ingesting Searching Browsing Sorting Metadata	There was no research data management platform for the faculty, therefore no policy guidelines, training advocacy, and stakeholder involvement	✗	The model does not either fully or partially cover this attribute.	✗	The model does not either fully or partially cover this attribute.
Policy Guidelines	Data Management Standards Compliance Security Monitoring		✓	It does not discuss RDM policies but draws attention to legal concerns that could hinder sharing and re-using of data.	✗	
Training and Advocacy	Data Creators Data Consumers Faculty Community		✓	It draws attention to the need for research skills in general, especially those related to cloud computing, visualizations, scientific techniques, experiments, data synopsis, classification, and corroboration	✗	
Stakeholders	Library Researchers Publishers Funders Research Office		✗	Emphasizes teamwork for the intensive production of research at the departmental, international, and cross-national levels but not on RDM	✗	
Recommended 5th Component: Role Players						
Sub-components			**Roles**			
			CCMF		**UCRDMF**	

Sub-component	Elements	Observations	✓/ ✗	Observations	✓/ ✗	Observations
Role Players	Data Creators	There was no research data management platform for the faculty, therefore no role players	✗	The model does not either fully or partially cover this attribute	✗	The model does not either fully or partially cover this attribute
	Data Consumers					
	Data Administrators					

The researcher's goal in conducting this study was to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology. However, the faculty research data repository platform's proposed model can be modified and applied in higher education institutions' faculties. The researcher also contends that adding to the body of knowledge in any field should indeed be done through shared understanding and lived experience to advance society. The design of the faculty research data repository platform for the UoT context in South Africa is thus best exemplified by the theoretical frameworks selected, literature review, meta-analysis review of research repositories in HEIs, methodology supported by pragmatism, results from the data collection instruments, interpretations, discussion of findings, and Delphi method.

This model is not meant to be authoritarian; it's a feature to expand the possibilities for developing faculty-level research data repository platforms within UoTs and the entire university system. The platform model for the faculty research data repository for the UoT context is presented in Figure 7.2.

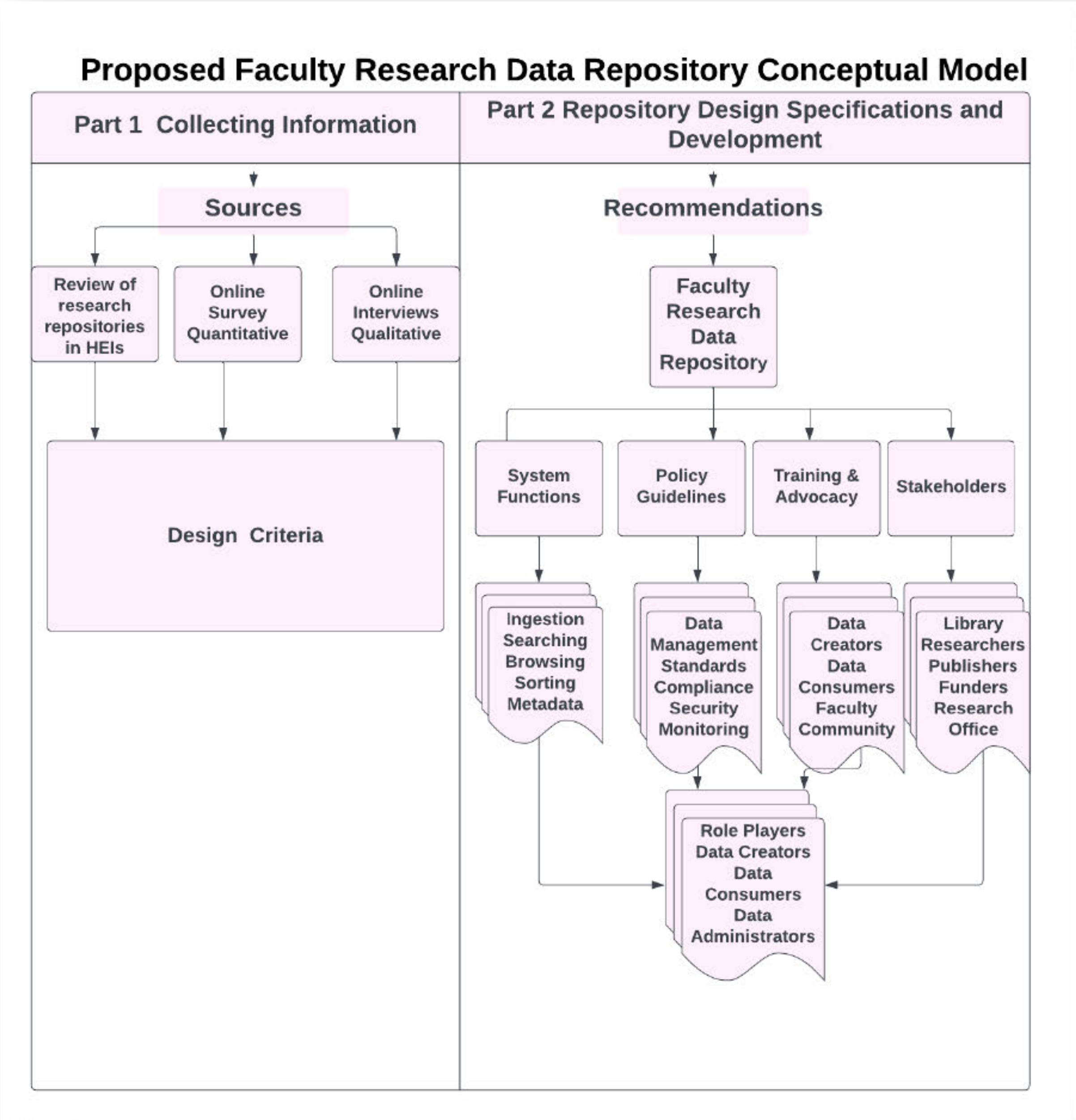

Figure 7.2. Proposed Faculty Research Data Repository Platform Model

The conceptual framework provides comprehensive instructions on planning and carrying out the Faculty Research Data Repository Platform design, as stated in section 7.3.4.2. The model will enable faculties within institutions to examine the demands of their students, the most recent data, where it is stored, and precisely who is in control. Based on their research and recommendations, they can design a platform for managing and disseminating this data in a regulated environment with other faculties and institutions.

7.5 Significant Contributions to the Body of Knowledge

A few significant contributions from this research have added to the comprehensiveness or complexity of the field's expertise.

(i) A concept model for a faculty research repository platform for a University of Technology

The research's main accomplishment was creating a simple RDM concept suitable for faculty research data repositories for the University of Technology based on the end-user requirements. The model's design may impact how policies are developed to oversee and enhance the potential for managing research data within the universities of technology faculties. The results of this study should open many further conversations about the creation, adoption, and implementation of similar research data management facilities in South African universities of technology and throughout Africa.

(ii) Contribution to the theory

The study adds significant literature on the sector of RDM, particularly when considering the setting of faculty repositories and the makeup of technology universities. The corroboration from thorough literature synthesis and analysis in chapters three and six indicates that this study represents the first substantial study to evaluate available research repositories and their advantages to propose a faculty repository set up informed by participants' recommendations. This was created for the faculty of a South African technology university. Over and above national boundaries, the study significantly contributes to the body of knowledge on RDM in Africa, where, previously, the majority of the literary works had been premised on proprietary platforms that were technically sophisticated and user-ready.

(iii)List of user requirements for the system architecture

One of the knowledge additions is a list of user requirements for a faculty research data repository, originally obtained from the participants' responses during the data collection phases and explicitly identifying their platform preferences. The description is also one that has not been created at this point and can serve as the foundation for more advancement.

7.6 Implications of the Study

The new evidence gathered from the research would be helpful to society in an array of ways. The goal of this study was to design a faculty research data repository platform that would be suitable for the University of Technology. It also examined research data management from a faculty perspective to support researchers. The study's conclusions may be helpful in South African and African universities' attempts to set up faculty repositories. By putting in place these faculty-associated repositories, difficulties related to a lack of facilities for managing research data could be overcome, particularly at universities of technology.

The requirement specifications for a faculty research data repository, compiled initially from the participants' responses during the data collection phases that clearly and unambiguously identify their ideal platforms, could be used as a model for developing faculty repositories in the long term and could be expanded upon.

When considering implementing faculty repositories and the makeup of universities of technology, the study significantly contributes to the literature on the RDM industry. Latent Dirichlet Allocation (LDA), a machine learning algorithm, was used in a meta-analysis of the research repositories at HEI to reveal critical themes consistent with the participants' results. This analysis might be helpful for many upcoming studies. The study's digital data-gathering strategy was effectively applied and displayed as an approach for other researchers facing similar challenges.

7.7 Future Research

A faculty at the Durban University of Technology was the subject of the current study's investigation into research data management practices. There are other faculties at the Durban University of Technology and other technology universities, though they weren't examined. Future research could concentrate on the faculties at provincial, national, or continental universities of technology.

Another study could go beyond the current user requirements by conducting a user evaluation of the proposed platform and making recommendations based on the research's findings.

7.8 Appraisal of the Chapter

Chapter Seven serves as a critical culmination of the entire investigation, succinctly revisiting the study's purpose and research objectives, providing an overview of the preceding chapters (Chapters One through Seven), and summarizing the key features of each chapter. It skillfully encapsulates the research journey, offering a comprehensive recap of the significant elements of the preceding chapters.

The chapter commences by revisiting the viewpoints on research data management, analysis of research data repositories, and the problem statement outlined in Chapter One. This reaffirms the grounding of the research in identified issues and sets the stage for the subsequent summarization of each chapter.

A pivotal aspect of the summary is the encapsulation of Chapter Two, where the study's guiding models—namely, the Data Audit Framework, the Community Capability Model Framework, and the User-Centered Research Data Management Framework—are reviewed along with the criteria for their selection. This is crucial for comprehending the methodological underpinnings and conceptual frameworks that shaped the research.

Chapter Three's literature review is highlighted for its exploration of critical concepts in the field, including a PRISMA-informed meta-analysis of research repositories in higher education institutions, underlining the scholarly context and background against which the current investigation operates.

The research design and data collection procedures, as outlined in Chapter Four, are also summarized, providing insight into the methodology and approaches employed in the study.

Chapters Five and Six, focusing on results and discussions, are summarized, emphasizing the outcomes obtained from both quantitative and qualitative data collection. This encapsulation showcases the research's analytical rigor and ability to derive meaningful insights from gathered data.

In conclusion, Chapter Seven encapsulates the significance of the study, summarizing essential advancements in the field of knowledge resulting from the investigation. It reiterates the value and relevance of the research and provides a concise yet holistic overview of the entire research journey, summarizing the chapters' core components and contributions to the academic domain.

REFERENCES

Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S., et al. 2021. Narratives and counter-narratives on data sharing in Africa, in *ACM FAccT Annual Conference* (Virtual Event Canada), 329–341. Available at https://doi.org/10.1145/3442188.3445897 (Accessed 23 March 2023).

Abbott, D. 2009. *DCC briefing papers: Introduction to curation*. Available at http://www.dcc.ac.uk/resources/briefingpapers/introductioncuration/Interoperability (Accessed 5 April 2021).

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. (Unpublished). Available: http://nectar.northampton.ac.uk/2736/ (Accessed 1 May 2021).

Ames, A. L. 2001. Users first! An introduction to usability and user-centered design and development for technical information and products. In: *IPCC 2001. Communication Dimensions. Proceedings IEEE International Professional Communication Conference* (Cat. No.01CH37271). IEEE. Available: https://doi.org/10.1109/ipcc.2001.971558 (Accessed 12 May 2021).

Armbruster, C. and Romary, L. 2010. Comparing repository types: challenges and barriers for subject-based repositories, research repositories, national repository systems, and institutional repositories in serving scholarly communication. *International Journal of Digital Library Systems*, 1(4): 62-63.

Anane-Sarpong, E., Wangmo, T., Ward, C. L., Sankoh, O., Tanner, M. and Elger, B. S. 2017. You cannot collect data using your own resources and put it on open access: perspectives from Africa about public health data-sharing. *Developing World Bioethic*s, 18(4): 1-12. Available: http://dx.doi.org/10.10.1111/dewb.12159 (Accessed 12 May 2022).

Anderl, R., Mecke, K., Klug, L. 2007. Advanced Prototyping With Parametric Prototypes. In: Cunha, P.F., Maropoulos, P.G. (eds) Digital Enterprise Technology. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-49864-5_59 (Accessed 13 May 2022).

Antell, K., Foote, J., Turner, J. and Shults, B. 2014. Dealing with data: science librarians' participation in data management at Association of Research Libraries Institutions. *College & Research Libraries, 75*(4): 557-574. Available: https://doi.org/10.5860/crl.75.4.557 (Accessed 20 May 2022).

Atkins, D. E. 2003. Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. Available:http://arizona.openrepository.com/arizona/handle/10150/106224 (Accessed 1 June 2021).

Ball, A. 2012. *Review of data management lifecycle models*. Bath, UK: University of Bath.

Babbie, E.R. 2016 *The practice of social research*. Fourteenth edition. Cengage Learning.

Bangani, S and Moyo, M. 2019. Data Sharing Practices among Researchers at South African Universities. Data Science Journal, 18(28): 1–14. DOI: https://doi.org/10.5334/dsj-2019-028 (Accessed 21 June 2022).

Barbieri, L., Angilica, A., Bruno, F. and Muzzupappa, M. 2013. Mixed prototyping with configurable physical archetype for usability evaluation of product interfaces. *Computers in Industry*, 64(3): 310–323.

Bellamy, C. 2012. Principle of methodology research design in social science. *Journal of Multidisciplinary Evaluation,* 8(18): 788-791.

Beran, T. N. and Violato, C. 2010. Structural equation modeling in medical research: a primer. *BMC Research,* 3: 267. Available: https://doi.org/10.1186/1756-0500-3-267 (Accessed 7 June 2021).

Berlin Declaration 2003. Berlin Declaration on open access to knowledge in the sciences and humanities. Available: http://oa.mpg.de/files/2010/04/Berlin-I-2.pdf (Accessed 13 June 2021).

Besançon, L., Peiffer-Smadja, N., Segalas, C., et al. 2021. Open science saves lives: lessons from the COVID-19 pandemic. BMC Med Res Methodol, 21, 117. https://doi.org/10.1186/s12874-021-01304-y (Accessed 21 June 2022).

Bond-Lamberty, B. 2018. Data sharing and scientific impact in eddy covariance research. *Journal of Geophysical Research: Biogeosciences,* 123(4): 1440-1443. Available: http://dx.doi.org/10.1002/2018JG004502 (Accessed 12 June 2021).

Borgman, C. L. 2015. *Big data, little data, no data: Scholarship in a networked world.* Cambridge, MA: The MIT Press.

Bowen, G.A. 2009. Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2): 27-40.

Bowler, L., Koshman, S., Oh, J. S., He, D., Callery, B. G., Bowker, G. and Cox, R. J. 2011. Issues in user-centered design in LIS. *Library Trends,* 59(4): 721-752.

Boyce, C. and Neale, P. 2006. *Conducting in-depth interviews: a guide for designing and conducting in-depth interviews for evaluation input*. Watertown, MA: Pathfinder International.

Buchenau, M. and Suri, J. F. 2000 Experience prototyping. In: Proceedings of the 3rd conference on designing interactive systems: processes, practices, methods, and techniques. 424–433. Available at https://doi.org/10.1145/347642.347802 (Accessed 12 June 2021).

Bugaje, M., and Chowdhury, G. 2017. Towards a More User-Centered Design of Research Data Management (RDM) Systems. In: Information: Interactions and Impact (i3).; 27-30 June; Aberdeen; 2017: 53-55.

Bugaje, M. I. 2019. A novel framework for user-centered research data management. PhD, Northumbria University.

Briney, K. A., Coates, H. and Goben, A. 2020. *Foundational practices of research data management.* Available: https://doi.org/10.3897/rio.6.e56508 (Assessed17 August 2021).

Brown, S., Bruce, R. and Kernohan, D. 2015. *Directions for research data management in UK Universities*. Available: http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf (Accessed 20 August 2021).

Brouwer, D. and Du Plessis, J. L. 2016. Occupational hygiene-related research in South Africa: development of a research repository. *Occupational Health Southern Africa,* 22(4): 34-36.

Bryman, A. 2006. Integrating quantitative and qualitative research: how is it done? *Qualitative Research,* 6(1): 97-113.

Burns, R. B. 2000. *Introduction to research methods*. London: Sage.

Byrne, D. 2017. *Methods: planning research, research questions*. Available: https://dx.doi.org/10.4135/9781526408525 (Accessed 1 September 2022).

Camburn, B., Viswanathan, V., Linsey, J., Anderson, D., Jensen, D., Crawford, R., Otto, K. and Wood, K. 2017. Design prototyping methods: state of the art in strategies, techniques, and guidelines. *Design Science*, 3: 1-33. Available: https://www.cambridge.org/core/services/aop-cambridgecore/content/view/560B306A5E799AEE54D30E0D2C1B7063/S2053470117000105a.pdf/design-prototyping-methods-state-of-the-art-in-strategies-techniques-and-guidelines.pdf (Accessed 1 September 2022).

Card, N. A. 2012. Literature review. In: Salkind, N. J. eds. *Encyclopedia of research design.* Thousand Oaks, CA: Sage Publications, 726-728. Available: http://dx.doi.org/10.4135/9781412961288 (Accessed 12 November 2021).

CARL Data Management Sub-Committee. 2009. *Research data: unseen opportunities*. Available: http://carl-abrc.ca/uploads/pdfs/data_mgt_toolkit.pdf (Accessed 20 August 2021).

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academ*y, *11*(2): 629-657. http://dx.doi.org/10.1353/pla.2011.0022 (Accessed 20 August 2021).

Castillo-Megchun, I., López-Rossell, C., Padilla-Rivera, M., Villalobos-Molina, R. and Tapia-Pancardo, D. 2021. Problems-based learning during COVID-19 pandemic: experiences by nursing students. *Open Journal of Nursing*, 11: 920-932. Available: 10.4236/ojn.2021.1111075 (Accessed 12 November 2021).

CCDA (ALCTS/CCS/Committee on Cataloging: Description and Access). 2000. *Task force on metadata: final report, June 16, 2000*. Available: <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html> (Accessed 1 September 2022).

Chau, P. Y. 1997. Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Sciences,* 28(2): 309–334.

Chawinga, D. W. 2019. Research data management in public universities in Malawi. (Doctoral thesis) University of the Western Cape, Cape Town. Available at http://hdl.handle.net/11394/6951 (Accessed 12 March 2021).

Chawinga, D. W. and Zinn, S. 2019. Global perspectives of research data sharing: a systematic literature review. *Library & Information Science Research,* 41 (2): 109-122. Available: https://doi.org/10.1016/j.lisr.2019.04.004. (Accessed 1 September 2022).

Chen, X. and Wu, M. 2017. Survey on the needs for chemistry research data management and sharing. *The Journal of Academic Librarianship*, 43(4): 346-353. Available: http://dx.doi.org/10.1016/j.acalib.2017.06.006 (Accessed 13 December 2022).

Chigwada, J., Chiparausha, B. and Kasiroori, J. 2017. Research data management in research institutions in Zimbabwe. *Data Science Journal,* 16(31): 1–9, DOI: Available: https://doi.org/10.5334/dsj-2017-03 (Accessed 13 December 2022).

Chiware, E. R. T. 2020. Open research data in African academic and research libraries: a literature analysis. Library Management, 41 (6/7): 383-399. Available: https://doi.org/10.1108/LM-02-2020-0027 (Accessed 13 December 2022).

Chiware, E. R. T., & Skelly, L. (2022). Open Science in Africa: What policymakers should consider. Frontiers in Research Metrics and Analytics, 7, 950139. https://10.3389/frma.2022.950139 (Accessed 13 December 2022).

Chiware, E.R. T. and Becker, D. A. 2018a. Research data management services in Southern Africa: a readiness survey of academic and research libraries. African Journal of Library, Archives and Information Science, 28 (1): 1-16. Available: https://www.ajol.info/index.php/ajlais/article/view/174148 (Accessed 13 December 2022).

Chiware, E. R. T. and Becker, D. A. 2018b. Research trends and collaborations by applied science researchers in South African universities of technology: 2007–2017. Journal of Academic Librarianship, 44(4): 468-476.

Chiware, E. R. T. and Mathe, Z. 2015. Academic libraries' role in research data management services: a South African perspective. South African Journal Libraries and Information Science. 81(2): 1-10.

Christensen, B. T. and Schunn, C. D. 2007. The relationship of analogical distance to analogical function and pre-inventive structure: the case of engineering design. *Memory & Cognition,* 35: 29–38.

Chowdhury, G. G. 2014. *Sustainability of scholarly information.* London: Facet Publishing.

Chowdhury, N., Katsikas, S. and Gkioulos, V. 2022. Modeling effective cybersecurity training frameworks: a Delphi method-based study. *Computers & Security,* 113: 1-15. Available: https://doi.org/10.1016/j.cose.2021.102551. (Accessed 10 January 2023).

Cidell, J. 2010. Content clouds as exploratory qualitative data analysis. *Area,* 42(4): 514-523.

Cimino, J. J., Ayres, E. J., Remennik, L., Rath, S., Freedman, R., Beri, A., Huser, V. 2014. The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): Design, contents, functionality and experience to date. *Journal of Biomedical Informatics*, 52:11–27. https://doi.org/10.1016/j.jbi.2013.11.004 (Accessed 12 April 2021).

Cohn, J. P. 2012. Dataone opens doors to scientists across disciplines. *BioScience, 62*(11): 1004-1004. Available: http://dx.doi.org/10.1525/bio.2012.62.11.16 (Accessed 10 January 2023).

Connaway, L. S. and Powell, R. R. 2004. *Basic research methods for librarians.* 10th ed. Greenwich, Conn: Libraries Unlimited.

Consultative Committee for Space Data Systems (CCSDS). 2012. *Reference model for an open archival information system (OAIS)*. Washington, DC: CCSDS Secretariat. Available: http://public.ccsds.org/publications/archive/650x0m2.pdf (Accessed 10 December 2021).

Corti, L., Van den Eynden, V., Bishop L. and Woollard, M. 2014. Managing and sharing research data: a guide to good practice. Sage.

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: http://doi.org/10.1177/0961000613492542.Abstract (Accessed 10 December 2021).

Creswell, J.W. 2018 *Designing and conducting mixed methods research*. Third edition. Edited by V.L. Plano Clark. Sage Publications.

Creswell, J. W. and Creswell, J. D. 2018. *Research design: qualitative, quantitative, and mixed methods approaches.* London: Sage Publications, Inc.

Critchlow, T. and Van Dam, K. K. 2016. *Data-intensive science*. Chapman and Hall/CRC.

Curdt, C. and Hoffmeister D. 2015. Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database. *Program: Electronic Library and Information Systems*, 49(4): 494-512. Available: https://doi.org/10.1108/PROG-02-2015-0016 (Accessed 5 December 2021).

Dai, S. Q., Li, H., Xiong, J., Ma, J., Guo, H. Q., Xiao, X. and Zhao, B. 2018. Assessing the extent and impact of online data sharing in eddy covariance flux research. *Journal of Geophysical Research: Biogeosciences*, *123*(1): 129-137. Available: http://dx.doi.org/10.1002/2017JG004277 (Accessed 5 December 2021).

Dahan, E. and Mendelson, H. 2001. An extreme-value model of concept testing. *Management Science,* 47(1): 102–116.

Deloitte, L. L. P. 2017. Assessing the value of TfL's open data and digital partnerships. Available: http://content.tfl.gov.uk/deloittereport-tfl-open-data.pdf. (Accessed 5 December 2022).

Denscombe, M. 2008. Communities of practice: a research paradigm for the mixed methods approach. *Journal of Mixed Methods Research*, 2(3): 270-283.

Denzin, N. K. and Lincoln, Y. S. 2005. Introduction: the discipline and practice of qualitative research. In: Denzin, N. K. and Lincoln, Y. S. eds. *The sage handbook of qualitative research*. 3rd ed. Thousand Oaks, CA: Sage, 1-32.

DePaolo, C. A. and Wilkinson, K. 2014. Using word clouds for analyzing qualitative assessment data. *Tech Trends,* 58(3): 38–44.

Dietrich, S., Ham, J. V., Pras, A., Rijswijk-Deij, R. V., Shou, D., Sperotto, A., Wynsberghe, A. V. and Zuck, L. D. 2014. Ethics in data sharing: developing a model for best practice. In: *2014 IEEE Security and Privacy Workshops*, 5-9. Available: https://doi.org/10.1109/SPW.2014.43 (Accessed 5 December 2022).

Dora, M. and Kumar, H. A. 2015. Managing Research Data in academic institutions: role of libraries, 2015. In CALIBER, 2015, Shimla, March 12-+14, 2015 [Conference paper]. [Online]. Available: http://eprints.rclis.org/24911/1/50.pdf (Accessed 5 December 2022).

Dow, S. P., Glassco, A., Kass, J., Schwarz, M. and Klemmer, S. R. 2009. *The effect of parallel prototyping on design performance, learning, and self-efficacy.* Stanford, CA: Stanford University HCI Group.

Drezner, J. A. 1992. *The nature and role of prototyping in weapon system development*. Santa Monica, CA: RAND. Available: https://doi.org/10.1017/dsj.2017.10 (Accessed 5 December 2022).

Du, Y. 2009. A review of structural equation modeling and its use in library and information studies. *Library & Information Science Research,* 31(4): 257–263.

Dulle, F. and Minish-Majanja, M. K. 2011. The suitability of the Unified Theory of Acceptance and Use of Technology (UTAUT) model in open access adoption studies. *Information Development,* 27(1): 32–45. Available: https://doi.org/10.1177/0266666910385375 (Accessed 23 January 2022)

Durban University of Technology. 2023. World University Rankings. Available at https://www.dut.ac.za/dut-ranked-top-5-in-south-africa-in-2023-world-university-rankings/ (Accessed April 2023).

Edinburgh University Data Library. 2011. Research data management handbook. 1-26. Available: http://www.docs.is.ed.ac.uk/docs/data-library/EUDL_RDM_Handbook.pdf (Accessed 23 January 2022).

Ekmekcioglu, C. and Rice, R. 2009. Edinburgh data audit implementation project: a final report. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8efce9b396aace31118d6ddf470c1c70dfee54ab (Accessed 13 March 2022).

Elsayed, A. M. and Saleh, E. I. 2018. Research data management and sharing among researchers in Arab universities: an exploratory study. *IFLA journal*, 44(4): 281-299. Available: http://dx.doi.org/10.1177/0340035218785196 (Accessed 23 January 2022).

Elsevier. 2019. *Sharing research data.* Available: www.elsevier.com/authors/author-services/research-data (Accessed 23 January 2022).

European Commission. 2012. *Scientific data: open access to research results will boost Europe's innovation capacity.* Available: http://europa.eu/rapid/press-release_MEMO-12-565_en.htm?locale=en (Accessed 13 March 2022)

European Commission. 2016. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020: Version 2.1. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (Accessed 20 April 2022).

Erichsen, J. F., Sjöman, H., Steinert, M. and Welo, T. 2021. Protobooth: gathering and analyzing data on prototyping in early-stage engineering design projects by digitally capturing physical prototypes. *AI EDAM*, 35(1): 65-80. Available: https://www.cambridge.org/core/services/aopcambridgecore/content/view/626E78A4E59E5D779BB5FC3F6009244B/S0890060420000414a.pdf/protobooth-gathering-and-analyzing-data-on-prototyping-in-early-stage-engineering-design-projects-by-digitally-capturing-physical-prototypes.pdf (Accessed 13 March 2022).

European Commission 2019. The EU's open science policy. Available at: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/openscience_en (accessed 18 June 2022).

European Parliament, Council. 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14. Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (5) (PDF) Data Interoperability. Available: https://inspire.ec.europa.eu/legislation-details/directive-20072ec-european-parliament-and-council (Accessed 13 March 2022).

Faundeen, J. 2010. Appraising U.S. geological survey science records. *Archival Issues,* 32(1): 7-22.

Fecher, B., Friesike, S. and Hebing, M. 2015. What drives academic data sharing? *PloS one,* 10(2): e0118053. Available: https://doi.org/10.1371/journal.pone.0118053 (Accessed 13 March 2022).

Fisher, R.A. 1920. A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. Mon. Notices Roy. Astron. Soc. 80 758–770; reprinted as Paper 12 in Fisher (1974).

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.

Frades, I. and Matthiesen, R. 2010. Overview on techniques in cluster analysis. In: Matthiesen, R. ed. *Bioinformatics Methods in Clinical Research. Methods in Molecular Biology*, Humana Press, 81–107. Available: https://doi.org/10.1007/978-1-60327-194-3_5 (Accessed 20 February 2021).

Gall, M. D., Gall, J. P. and Borg, W. R. 2006. *Educational research: an introduction*. 8th ed. New York: Longman.

Gallotta, B., Garza-Reyes, J.A. and Anosike, A. 2018. Using the Delphi method to verify a framework to implement sustainability initiatives: *Proceedings of the International Conference on Industrial Engineering and Operations Management*. Bandung, Indonesia, 6-8 March 2018, IEOM Society, 231-241. Available: http://hdl.handle.net/10545/622171 (Accessed 12 October 2022).

Gelo, O. Braakmann, D. and Benetka, G. 2008. Quantitative and qualitative research: beyond the debate. *Integrated Psychology Behavior,* 42: 266-290.

George, D., & Mallery, P. 2003. *SPSS for Windows step by step*: A simple guide and reference. 11.0 update (4th ed.). Boston, MA: Allyn & Bacon.

Gerber, E. and Carroll, M. 2012. The psychological experience of prototyping. *Design Studies,* 33, 64–84.

Gero, J. S. and Lindemann, U. 2005. *Human Behaviour in Design 05*. Key Centre of Design Comp & Cogntn.

Gill, C., Sanders, E. and Shim, S. 2011. Prototypes as inquiry, visualization, and communication. In: The *International Conference on Engineering and Product Design Education*. London, UK.

Golafshani, N. 2003. Understanding reliability and validity in qualitative research. *The Qualitative Report,* 8(4): 597-607.

Gonzalez, A. and Peres-Neto, P.R. 2015. Act to staunch loss of research data. *Nature*, 520: 436. Available: https://doi.org/10.1038/520436c (Accessed 12 October 2022).

Gordon, V. S. and Bieman, J. M. 1995. Rapid prototyping: lessons learned. *IEEE Software,*12(1): 85–95.

Grabowski, B. 2016. P< 0.05" might not mean what you think: American Statistical Association clarifies P values. *JNCI: Journal of the National Cancer Institute*, 108(8): 4-5.

Grant, C. and Osanloo, A. 2014. Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house." *Administrative Issues Journal*, 4(2): 12-26. Available: http://dx.doi.org/10.5929/2014.4.2.9 (Accessed 12 October 2022).

Gray, D.E. 2009. *Doing social research in the real world*. Los Angeles: Sage.

Gray, J.R., Grove, S.K. and Sutherland, S. 2016. *Burns and Grove's the practice of nursing research- Appraisal, synthesis, and generation of evidence*. Elsevier Health Sciences.

Green, G. and Smrcek, L. 2006. On the developing role of physical models in engineering design education. *European Journal of Engineering Education*, 31: 191–200.

Guba, E. and Lincoln, Y. 1994. Competing paradigms in qualitative research. In: Denzin, N. K. and Lincoln, Y. S. eds. *Handbook of qualitative research*. Thousand Oaks, California: Sage Publishers, 105–117

Guedon, J.C. 2015. *Open data and science: Towards optimizing the research process*. Available: https://www.dataone.org/webinars/open-data-and-science-towards-optimizing-research-process (Accessed 17 May 2021)

Hair, J. F., Hult, G. T. M., Ringle, C. and Sarstedt, M. 2016. *A primer on partial least squares structural equation modeling*. (PLS-SEM). Sage Publications.

Heath, T. and Bizer, C. 2011. *Linked data: evolving the web into a global data space*. Morgan & Claypool.

Heffetz, O. and Ligett, K. 2014. Privacy and data-based research. *Journal of Economic Perspectives*, 28 (2): 75-98.

Hess, D. R. 2004. Retrospective studies and chart reviews. Respiratory Care, 49(10), 1171-1174. PMID: 15447798.

Hess, T. and Summers, J. D. 2013. Case study: evidence of prototyping roles in conceptual design. In*: DS 75-1: Proceedings of the 19th International Conference on Engineering Design (ICED13)*, Design for Harmonies: Design Processes, Seoul, Korea, Clemson University, 249-258.

Higgins, S. 2008. The DCC curation lifecycle model. *International Journal of Digital Curation*, *3*(1): 134-140. Available: http://dx.doi.org/10.2218/ijdc.v3i1.48 (Accessed 11 June 2021).

Higgins, S. 2011. Digital curation: the emergence of a new discipline. *The International Journal of Digital Curation*, *2*(6): 78-88. Available: http://dx.doi.org/10.2218/ijdc.v6i2.191 (Accessed 11 June 2021).

Hlatshwako, T.G. et al. 2021. Online health survey research during COVID-19 The Lancet, 2(3):76-77. Available at https://doi.org/10.1016/S2589-7500(21)00002-9 (Accessed 11 June 2022).

Horizon 2020 European Research Council. 2016. Guidelines on FAIR Data Management in Horizon 2020. 1-12. Available: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 11 June 2021).

International Organization for Standardization. 1999. ISO/IEC 13407: Human-Centered Design Processes for Interactive Systems. Available: ISO 13407:1999(en), Human-centred design processes for interactive systems (Accessed 11 July 2021).

Israel, G. D. 2013. Determining sample size. University of Florida: Institute of Food and Agricultural Sciences.

Ivankova N.V., Creswell, J.W. and Stick, S.L. 2006. Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1): 3–20. Available at https://10.1177/1525822X05282260 (Accessed 12 April 2022).

Jahnke, L. M. and Asher, A. 2012. The problem of data: data management and curation practices among university researchers. In: L. Jahnke, A. Asher and S. Keralis. *The Problem of Data*. Washington, DC: Council on Library and Information Resources. 1-48. Available: https://www.clir.org/pubs/reports/pub154/pub154.pdf (Accessed 11 July 2021).

Jang, J. and Schunn, C. D. 2012. Physical design tools support and hinder innovative engineering design. *Journal of Mechanical Design.* 134, 041001.

Jeng, W., He, D. and Chi, Y. 2017. Social science data repositories in data deluge: A case study of ICPSR's workflow and practices. *The Electronic Library*, 35(4): 626-649. Available: https://doi.org/10.1108/EL-11-2016-0243 (Accessed 25 May 2022).

Jeng, W. and He, D. 2022, Surveying research data-sharing practices in US social sciences: a knowledge infrastructure-inspired conceptual framework, *Online Information Review*, 46(7):1275-1292. https://doi.org/10.1108/OIR-03-2020-0079 (Accessed January 2023).

Jernigan, T. L., Brown, T. T., Hagler, D. J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W. K., Bloss, C. S., Murray, S. S., Schork, N. and Kennedy, D. N. 2016. The Paediatric Imaging, Neurocognition, and Genetics (PING) data repository. *Neuroimage*, 124: 1149-1154.

Jensen, L. S., Özkil, A. G., & Mortensen, N. H. 2016. Prototypes in engineering design: Definitions and strategies. In International Design Conference - DESIGN 2016, Dubrovnik, Croatia.

Johnson, R. B. and Onwuegbuzie, A. J. 2004. Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, 33: 14-26.

Johnson, B., and Shneiderman, B. 1991. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91,* San Diego, CA, USA. 284-291. Available at https://ieeexplore.ieee.org/document/175815 (Accessed 20 April 2022)

Jones, S., Ross, S., Ruusalepp, R. and Dobreva, M. 2009. Data Audit Framework Methodology, draft for discussion, version 1.8. Glasgow, HATII, 1-71. Available at https://10.13140/RG.2.1.3389.0969 (Accessed 20 April 2021).

Kabanda, S. M., Cengiz, N., Rajaratnam, K., Watson, B. W., Brown, Q., Esterhuizen, T. M. and Moodley, K. 2023. Data sharing and data governance in sub-Saharan Africa: Perspectives from researchers and scientists engaged in data-intensive research, *South African Journal of Science*, 119(5/6). https://10.17159/sajs.2023/15129 (Accessed 5 April 2023).

Kahn, M., Higgs, R., Davidson, J., and Jones, S. 2014. Research data management in South Africa: how we shape up. *Australian Academic and Research Libraries*, 45(4): 296-308. Available at https://10.1080/00048623.2014.951910 (Accessed 20 March 2021).

Kallet, R. H. 2004. How to write the methods section of a research paper. *Respiratory Care*, 49: 1229-1232.

Kanngieser, A., Neilson, B. and Rossiter, N. 2014, What is a research platform? Mapping methods, mobilities and subjectivities. *Media, Culture and Society*, 36 (3): 302-318.

Katabalwa, A.S, Bates, J., and Abbott, P. 2021. Potential opportunities and risks of sharing agricultural research data in Tanzania, IASSIST Quarterly 45(3-4): 1-8. Available at https://doi.org/10.29173/iq997 (Accessed 14 April 2022).

Kem¨Eny, Z., Beregia, J.R., Erdos, G. and Nacsa, J. 2016, The MTA SZTAKI smart factory: platform for research and project-oriented skill development in higher education. *Procedia CIRP*. 54: 53-58.

Kennan, M., Corrall, S. and Afzal, W. 2014. Making space in practice and education: Research support services in academic. *Library Management*, 35: 666-683. Available: http://doi.org/10.1108/LM-03-2014-0037 (Accessed 30 November 2021).

Kershaw, T. C., Hölttä-Otto, K. and Lee, Y. S. 2011. The effect of prototyping and critical feedback on fixation in engineering design. In: 33rd *Annual Meeting of the Cognitive Science Society* 2011, Boston, MA, USA. Available at https://escholarship.org/uc/item/5kq942pr (Accessed 30 November 2021).

Kirshenblatt-Gimblett, B. 2006. Part 1. What is research design? The context of design performance studies methods course syllabus. In: W. M. K. Trochim (Ed.), *Research methods knowledge base*. New York: New York University.

Knight, G. 2015. Building a research data management service for the London School of Hygiene & Tropical Medicine. *Program: Electronic Library and Information Systems*, 49(4): 424-439. Available: https://doi.org/10.1108/PROG-01-2015-0011 (Accessed 10 November 2022).

Knottnerus J.A. 2016. Research data as a global public good. *Journal of Clinical Epidemiology,* 70:270-1. Available at http://10.1016/j.jclinepi.2015.05.034 (Accessed 15 October 2021).

Koltay, T. 2017. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1): 3–14. Available at https://doi.org/10.1177/0961000615616450 (Accessed 18 December 2021).

Koopman, M. M. and de Jager, K. 2016. Archiving South African digital research data: How ready are we? *South African Journal of Science*. 112 (7/8): 1-7. Available: http://dx.doi.org/10.17159/sajs.2016/20150316 (Accessed 12 December 2022).

Kouper, I., Konkiel, S. R., Liss, J. A. and Hardesty, J. L. 2013. Collaborate, automate, prepare, prioritize: creating metadata for legacy research data. In: *Proceedings of the 2013 International Conference on Dublin Core and Metadata Applications (DCMI'13)*, Muriel Foulonneau and Kai Eckert (Eds.). Dublin Core Metadata Initiative: 41-46.

Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, Koczan P, Knight P, Marsolier C, McKendry RA, Ross E, Sasse A, Sullivan R, Chaytor S, Stevenson O, Velho R, and Tooke J. 2016. Who Owns the Data? Open Data for Healthcare. *Frontiers in Public Health*, 17(4):7. Available at https://10.3389/fpubh.2016.00007 (Accessed 29 March 2021).

Kothari, C.R. 2004. *Research methodology: Methods and techniques*. 2nd ed. New Delhi: New Age International Publishers.

Krier, L. and Strasser, C. A. 2014. *Data management for libraries: A LITA guide*. Chicago: American Library Association.

Kumar, S. 2010. Undergraduate perceptions of the usefulness of Web 2.0 in higher education: survey development. Available: http://plaza.ufl.edu/swapnak/ecel09Kumar.pdf (Accessed 10 November 2022).

Kurvinen, E., Koskinen, I. and Battarbee, K. 2008. Prototyping social interaction. *Design*, 24: 46–57.

Kvale, S., and Brinkmann, S. 2009. *Interviews: Learning the craft of qualitative research interviewing*. Beverly Hills, California: Sage Publications.

Lagoze, C, Van de Sompel, H. 2008. ORE user guide—Primer: Technical report, Open Archives Initiative. Object Reuse and Exchange. Available at http://www.openarchives.org/ore/0.9/primer.html (Accessed 10 November 2022).

Lather, P. 1986. Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange,* 17: 63–84. Available: https://doi.org/10.1007/BF01807017 (Accessed 10 November 2022).

Lazar, J. 2006. *Web usability: A user-centered design approach*. Boston: Pearson, Addison Wesley.

Leedy, P. D. and Omrod, J. E. 2010. *Practical research: planning and design*. 9th ed. Upper Saddle River, NJ: Pearson Education.

Leng, C. B., Ali, K. M. and Hoo, C. E. 2016. Open access repositories on open educational resources: feasibility of adopting the Japanese model for academic libraries. *Asian Association of Open Universities Journal*, 11 (1): 35-49.

Lennings, A. F., Broek, J. J., Horváth, I., Sleijffers, W. and Smit, A. D. 2000. Editable physical models for conceptual design. Third international symposium on tools and methods of competitive engineering. Delft University Press, 665-674.

Limani, F., Hajra, A., Ferati, M., and Radevski, V. 2020. Requirements and recommendations for university research data repository: A case study, In: *Proceedings of the 18th International Conference e-Society*, April 2-4: 51-58. IADIS Press. Available: https://www.diva-portal.org/smash/get/diva2:1449234/FULLTEXT01.pdf (Accessed 2 March 2022).

Lyon, L. 2012. The informatics transform: Re-engineering libraries of the data decade. *International Journal of Digital Curation*, *7*(1): 126-138. Available: https://doi.org/10.2218/ijdc.v7i1.220 (Accessed 2 March 2022).

Lyon, L., Patel, M. and Takeda, K. 2014. Assessing requirements for research data management support in academic libraries: Introducing a new multi-faceted capability tool. Available: http://ozk.unizd.hr/proceedings/index.php/lida/article/view/121 (Accessed 11 January 2021).

Maciaszek, L. 2007. *Requirements Analysis and Systems Design*. 3rd ed. New Jersey: Pearson Education UK.

Mackenzie, N. and Knipe, S. 2006. Research dilemmas: paradigms, methods, and methodology. *Issues in Educational Research*, 16(1): 193-205. Available at http://www.iier.org.au/iier16/mackenzie.html (Accessed 12 August 2021).

Majyambere, M. 2014. Information seeking behaviour of humanities/arts international postgraduate students in public universities in KwaZulu-Natal province, South Africa. Doctoral thesis) University of Kwa-Zulu Natal, Durban, South Africa. Available at http://hdl.handle.net/10413/12111

Mannheimer, S., Sterman, L. and Borda, S. 2016. Discovery and reuse of open datasets: An exploratory study. *Journal of eScience Librarianship*. Available: https://doi.org/10.7191/jeslib.2016.1091 (Accessed 11 January 2022).

Manu, T. R., Viral, A., Madesh, G., Shashikumara, A. A., Panna, C. and Prasanna, K. 2018. Analysis of research data repositories in India. In: *International Conference on Knowledge Organisation in Academic Libraries (I-KOAL-2018).* Library Professionals Association (LPA) and University of Hyderabad, Hyderabad.

Martines-Uribe, L. 2009. Using the data audit framework: An Oxford case study scoping digital repository services for research data management. Available at http://www.disc-uk.org/docs/DAF-Oxford.pdf (Accessed 13 August 2021).

Martone, M. E, Garcia-Castro, A and Van den Bos, G. R. 2018. Data sharing in psychology. *American Psychologist*, 73(2): 111–125. Available at https://doi.org/10.1037/amp0000242 (Accessed 11 January 2022).

Matlatse, R. L. 2016. An evaluation of a structured training event aimed at enhancing the Research Data Management (RDM) knowledge and skills of library and information science (LIS) professionals in South African higher education institutions (HEIs) (Unpublished master's thesis). MA thesis, University of Pretoria, South Africa.

Maxwell, J. A. 2012. *Qualitative research design: An interactive approach*. 3rd ed. London: Sage Publications.

McEwan, B., 2018. Literature Review, In Allen, M. (ed), *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications. Available at https://dx.doi.org/10.4135/9781483381411 (Accessed 11 January 2022).

Mertens, D. M. 1998. *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches*. London: Sage Publications.

Mi, X., Bernardy, R. and Schmidt, L. 2021. Building an archaeological data repository: a digital library and digital humanities collaboration at the University of South Florida. *International Journal on Digital Libraries*, 22 (1): 135-145.

Mischo, W., Schlembach, M. and O'Donnell, M. 2014. An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals. *Journal of eScience Librarianship*, 3(1): 31-43. Available: https://doi.org/10.7191/jeslib.2014.1060 (Accessed 7 May 2021).

Mwelwa, J., Boulton, G., Wafula, J. M., and Loucoubar, C. 2020. Developing open science in Africa: barriers, solutions and opportunities. *Data Science. Journal*. 19(31). Available at https://10.5334/dsj-2020-031 (Accessed 23 March 2023).

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. 2010, Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement. *International Journal of Surgery*, 8 (5): 336-341.

Morgan, D. L. 2007. Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1): 48-76. Available: http://dx.doi.org/10.1177/2345678906292462 (Accessed 7 May 2021).

Muchaonyerwa, N. 2015. Knowledge sharing strategies in university libraries in Kwa-Zulu Natal province of South Africa. Doctoral thesis) University of Kwa-Zulu Natal, Durban South Africa. Available at http://handle.net/10413/13085 (Accessed 12 April 2021).

Mueller, R. O. 1996. *Basic principles of structural equation modeling*. New York: Springer.

Musgrave, S., 2003. The metadata dynamic: ensuring data has a long and healthy life. In *Proceedings of the Fourth ASC International Conference*. 119-129.

Mutsvunguma, G. 2019. Institutional repositories as platforms for open access in South African universities: the case of University of KwaZulu-Natal (Doctoral thesis). University of Kwa-

Zulu Natal, Durban, South Africa. Available at: https://researchspace.ukzn.ac.za/handle/10413/16499 (Accessed 12 April 2021).

Myung, I. J. 2003. Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47(1): 90-100. Available at. https://doi.org/10.1016/S0022-2496(02)00028-7. (Accessed 5 January 2023)

National Science Board. 2005. Long-lived digital data collections: Enabling research and education in the 21st Century. Available: http://www.nsf.gov/pubs/2005/nsb0540/ (Accessed 4 April 2022).

NASA Socioeconomic Data and Applications Center (SEDAC) Long-Term Archive. 2010. Appraisal for accession to the SEDAC LTA. Available: http://sedac.ciesin.columbia.edu/lta/Appraisal.html (Accessed 4 April 2022).

Nhendodzashe, N. 2017. Exploring the feasibility of offering research data management services at the University of Zimbabwe Library (Unpublished) Masters thesis, National University of Science and Technology, Bulawayo.

Ndhlovu, U. P. and Ngwenya, S. 2017. Research Data Management Services: An investigation of research data management practices at the National University of Science and Technology (NUST). 7th Annual International Conference on Communication and Information Science, Caribbea Bay, Kariba, Zimbabwe, 21- 24 August 2017.

Ng'eno, E. J. 2018. Research data management in Kenya's agricultural research institutes. (Doctoral thesis), University of KwaZulu-Natal, KwaZulu-Natal, South Africa. Available at http://hdl.handle.net/10413/15908 (Accessed 21 March 2021).

NISO (National Information Standards Organization). 2004. Understanding metadata. Bethesda, MD: NISO Press. Available: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf (Accessed 3 January 2021).

Noko, P., and Ngulube, P. 2015. A vital feedback loop in educating and training archival professionals: a tracer study of records and archives management graduates in Zimbabwe. Information Development, 31(3): 270-283. Available: https://doi.org/10.1177/0266666913510308 (Accessed 10 February 2022).

Nugroho, R. P., Zuiderwijk A., Janssen M., and de Jong M. 2015. A comparison of national open data policies: lessons learned. Transforming Government: People, Process and Policy, 9(3): 286–308.

Office of Scholarly Communication. 2021. University of Cambridge Research Data Management Policy Framework. Available: https://doi.org/10.17863/CAM.10528 (Accessed 3 January 2021).

Office of Science and Technology Policy. 2013. USA Memorandum: Increasing access to the results of federally funded scientific research. 1-51. Available: https://sgp.fas.org/crs/misc/RL34736.pdf (Accessed 3 January 2021).

Ohaji, I. K. 2016. Research data management: An exploration of the data librarian role in New Zealand research organisations. PhD, Victoria University of Wellington, Australia. Available: http://researcharchive.vuw.ac.nz/handle/10063/5657 (Accessed 1 July 2022).

Olayiwola, B., Akinwande V., Oluwatoyosi J. and Ahmed I. 2016. Open data portals in Africa: an analysis of open government data initiatives. *African Journal of Library, Archives and Information Science,* 26(2): 97-106.

Olivier A. 2016. The Impact of a Move Towards Open Data in West Africa originally published as The Impact of a Move Towards Open Data in West Africa. Available: https://alais.org/the-impactofamovetowardsopendatainwestafrica/#:~:text=The%20Argument%20for%20Open%20Data,becomes%20more%20transparent%20and%20efficient. (Accessed 1 July 2022).

O'Neill, B., Ryan, C., Roy, S. and Simes, T. 2020. Supporting nursing faculty with a digital repository of simulation resources. *Teaching and Learning in Nursing,* 15 (3): 175-180.

Onwuegbuzie, A. J. and Collins, K. M. 2007. A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2): 281-316.

Onwuegbuzie, A. J. and Leech, N. L. 2007. A call for qualitative power analyses. *Quality & Quantity*, 41(1): 105-121. Available: http://dx.doi.org/10.1007/s11135-005-1098-1 (Accessed 1 July 2022).

Otto, K. and Wood, K. 2001. *Product design: Techniques in reverse engineering and new product design*. Prentice-Hall.

Patel, D. 2016. Research data management: a conceptual framework. *Library Review*. 65 (4/5): 226-241.

Patterton, L. H. 2016. Research data management practices of emerging researchers at a South African Research Council, MIS, University of Pretoria, Pretoria, Available: http://hdl.handle.net/2263/59502 (Accessed 3 January 2023).

Patterton, L., Bothma, T. J., and van Deventer, M. J. 2018. From planning to practice: an action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science*, 84(2): 14-26.

Pearlman, R. 2009. NASA erased first moonwalk tapes, but restores copies. Available: https://www.space.com/6994-nasa-erased-moonwalk-tapes-restores-copies.html (Accessed 3 January 2023).

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11): 559-572.

Perakyla, A. and Ruusuvuori, J. 2011. Analyzing talk and text. In Denzin, N.K. and Lincoln, Y.S. ed. *The Sage handbook of qualitative research*. California: Sage Publications.

Pienaar, H. 2011. Survey of research data management practices at the University of Pretoria: undertaken by the Department of Library Services to improve research practices at the University.Available:https://repository.up.ac.za/bitstream/handle/2263/17475/Pienaar_Survey(2011).pdf?sequence=1 (Accessed 3 November 2022).

Pickard, A. J. 2007. *Research methods in information.* London: Facet Publishing.

Polit, D.F. and Beck, C. T. 2004. *Nursing research: principles and methods*. Philadelphia: Lippincott Williams and Wilkins.

Pouchard, L. (2015). 'Revisiting the Data Lifecycle with Big Data Curation.' International Journal of Digital Curation, 10(2): 342. Available at https://10.2218/ijdc.v10i2.342 (Accessed 28 August 2022).

Qi, X., Deng, H. and Guo, X. 2017. Characteristics of retractions related to faked peer reviews: an overview. *Postgraduate Medical Journal*, 93 (1102): 499–503. Available: https://pmj.bmj.com/content/93/1102/499 (Accessed 3 November 2022).

Qin, J., Crowston, K., & Lyon, L. (2016). Building capabilities for sustainable research data.

Management Practices. https://www.asist.org/files/meetings/am16/Building_Capabilities.pdf (Accessed 10 December 2012)

Queensland University of Technology. UQ policy and procedures library. Management of Research Data Policy. 2016. Available: https://ppl.app.uq.edu.au/content/4.20.06-research-data-management#Policy (Accessed 3 November 2022).

Raszewski, R., Goben, A. H., Bergren, M. D., Jones, K., Ryan, C., Steffen, A. D. and Vonderheid, S.C. 2021. A survey of current practices in data management education in nursing doctoral programs. *Journal of Professional Nursing*, 37 (1): 155-162.

Ravitch, S. M. and Riggan, M. 2016. *Reason & rigor: How conceptual frameworks guide research*. London: Sage Publications.

Ray, J. M. 2014. *Research data management: Practical strategies for information professionals.* West Lafayette, Indiana: Purdue University Press.

Re3data.org, Registry of Research Data Repositories. 2022. Available: https://doi.org/10.17616/R3D (Accessed 6 July 2022).

Republic of South Africa, Department of Higher Education and Training (DHET). 2016. Report on the evaluation of the 2015 universities' research outputs: Evaluated in terms of the policy and procedures for the measurement of research output of public higher education institutions. Pretoria: DHET.

Republic of South Africa, Department of Higher Education and Training (DHET). 2022. Report on the evaluation of the 2020 universities' research outputs: Evaluated in terms of the policy and procedures for the measurement of research output of public higher education institutions. Pretoria: DHET.

Research Information Network. 2008. Stewardship of digital research data: a framework of principles and guidelines. Available at http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital- research-data-principles-and-guidelines (Accessed December 5, 2022).

Rockinson-Szapkiw, A. J. and Spaulding, L. S. 2014. *Navigating the doctoral journey: a handbook of strategies for success*. Lanham, MD: Rowman and Littlefield Publishers.

Ross-Hellauer, T. (2017). Open peer review: Bringing transparency, accountability, and inclusivity to the peer review process. Impact of Social Sciences Blog. Retrieved April 23, 2022, from http://eprints.lse.ac.uk/85029/1/impactofsocialsciences-2017-09-13-open-peer-review-bringing.pdf

Rossman, G. B. and Wilson, B. L. 1985. Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, 9(5): 627-643.

Rothenberg, J. 1990. Prototyping as Modeling: What is being modeled. RAND Corp. Saakes, D. 2009 Big lampan lamps: designing for DIY. In: *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, pp. 403–404.

Rousi, A. M. 2023. Using current research information systems to investigate data acquisition and data sharing practices of computer scientists. *Journal of Librarianship and Information Science*, *55*(3), 596–608. https://doi.org/10.1177/09610006221093049 (Accessed 5 April 2023).

Rowhani-Farid, A. and Barnett, A. G. 2016. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ Open*, 6(10): e011784. Available: http://dx.doi.org/10.1136/bmjopen-2016-011784 (Accessed 10 October 2021).

Saib, M. O., Rajkoomar, M., Naicker, N. and Olugbara, C.T. 2022. Digital pedagogies for librarians in higher education: a systematic review of the literature. *Information Discovery and Delivery,* 51(1): 13-25.

Sale, A. 2006. Comparison of IR content policies in Australia. *First Monday*, 11(4). Available: http://eprints.utas.edu.au/264 (Accessed 10 October 2021).

Sam, C., Naicker, N. and Rajkoomar, M. 2020. Meta-analysis of artificial intelligence works in ubiquitous learning environments and technologies. *International Journal of Advanced Computer Science and Applications,* 11(9). Available at http://dx.doi.org/10.14569/IJACSA.2020.0110971 (Accessed 12 August 2021).

Sansone, S.A *et al*. 2020. Data Repository Selection: Criteria That Matter. Available at https://zenodo.org/record/4084763

Sardanelli, F., Alì, M., Hunink, M.G., Houssami, N., Sconfienza, L.M. and Di Leo, G. 2018. To share or not to share? Expected pros and cons of data sharing in radiological research. *European Radiology,* 28: 2328–2335. Available at https://doi.org/10.1007/s00330-017-5165-5 (Accessed 12 December 2021).

Sarstedt, M. and Mooi, E. 2019. Regression Analysis. In: *A Concise Guide to Market Research. Springer Texts in Business and Economics.* Springer, Berlin, Heidelberg. Available: https://doi.org/10.1007/978-3-662-56707-4_7 (Accessed 10 October 2021).

Sarter, M. 2006. The consequences of atheoretical, task-driven experimentation: Theoretical comment on Paban, Chambon, Jaffard, and Alescio-Lavtier (2005). *Behavioral Neuroscience*, 120(2): 493-494.

Satzinger, J., Burd, S. and Jackson, R. 2016. *Systems analysis and design in a changing world*. 7th ed. Australia: Cengage Learning.

Saunders, M., Lewis, P. and Thornhill, A. 2012. *Research methods for business students*. 6th ed. London: Prentice Hall.

Schöpfel, J., Ferrant, C., André, F. and Fabre, R. 2018. Research data management in the French National Research Center (CNRS). *Data Technologies and Applications*, 52(2): 248-265. Available: http://dx.doi.org/110.1108/DTA-01-2017-0005 (Accessed 3 September 2022).

Schofield, W. 2006. Survey sampling. In: Sapford, R. and Jupp, R. *Data collection and analysis*. 2nd ed. Los Angelos: Sage Publications, 26-55.

Schrage, M. 1993. The culture(s) of prototyping. *Design Management Journal (Former Series)* 4: 55–65.

Schubert, T. F., Jacobitz, F. G., Morse, M. S. and Ngo, T. T. 2012. The impact of a prototype exemplar on design creativity: a case study on novice designers. *In The ASEE Annual Conference and Exposition*, San Antonio, TX. Available at https://strategy.asee.org/22064 (Accessed 12 November 2021).

Schumacher, J., and Vande Creek, D. 2015. Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation*, 10(2): 96-109. Available: http://dx.doi.org/10.2218/ijdc.v10i2.32 (Accessed 3 September 2022).

Schumacker, R. E., and Lomax, R. G. 2004. *A beginner's guide to structural equation modelling*. 2nd ed. Mahwah, NJ: Erlbaum

Scott, M. 2014. Research data management (PhD). University of Southampton, United Kingdom. Available: http://dx.doi.org/10.5258/SOTON/374711 (Accessed 30 June 2021).

SDMX Initiative (n.d.) SDMX - statistical data and metadata exchange. Available: https://sdmx.org/ (Accessed 3 September 2022).

Sejane, L. 2017. Access to and use of electronic information resources in the academic libraries of the Lesotho Library Consortium. (Doctoral thesis) University of Kwa-Zulu Natal, Durban South Africa. Available at http://hdl.handle.net/10413/14345 (Accessed 13 March 2022).

Sekaran, U., and Bougie, R. 2010. *Research methods for business: A skill building approach.* 5th ed. Chichester, U.K.: John Wiley and Sons.

Shakeri, S. 2013. Data curation perspectives and practices of researchers at Kent State University's Liquid Crystal Institute: A case study (PhD). Kent State University, USA. Available: http://rave.ohiolink.edu/etdc/view?acc_num=kent1385382943 (Accessed 3 September 2022).

Shearer, K., Haigh, S. and Whitehead, M. 2015. Portage: Supporting Canadian innovation through shared expertise and stewardship of research data. IFLA WLIC 2015 - Cape Town, South Africa. Available: http://library.ifla.org/id/eprint/1263/1/198-shearer-en.pdf (Accessed 30 June 2021).

Sheenan, J. 2016. *Increasing Access to the Results of Federally Funded Science*. White House: Bill Announcement.

Shen, Y. 2015. Research data sharing and reuse practices of academic faculty researchers: A study of the Virginia Tech data landscape. *International Journal of Digital Curation*, 10(2): 157-175. Available: http://dx.doi.org/10.2218/ijdc.v10i2.359 (Accessed 1 July 2021).

South Africa. 2013. The Protection of Personal Information Act, No. 4 of 2013. Available at https://www.gov.za/documents/protection-personal-information-act (Accessed 12September 2022).

South Africa. 2019. White Paper on Science, Technology, and Innovation. Available: https://www.dst.gov.za/images/2019/FINAL-White-Paper-to-Cabinet_11-March-2019.pdf (Accessed 1 July 2021).

South African Council on Higher Education (SACHE) 2010. Universities of technology:Deepening the debate. Kagisano, 7. Available: http://www.che.ac.za/media_and_publications/kagisano-series/kagisano-issue-number-7universities-technologydeepening (Accessed 1 July 2021).

Southeastern Universities Research Association. 2012. Model Language for Research Data Management Policies ASERL / SURA Research Data Coordinating Committee. 1-6. Available: http://www.aserl.org/wpcontent/uploads/2013/01/ASERLSURA_Model_Language_RDM_Policy_Language_FINAL.pdf (Accessed 18 August 2021).

Stanley, L. M. and Edwards, M.C. 2016. Reliability and Model Fit. *Educational and Psychological Measurement,* 76(6): 976-985. Available at http://10.1177/0013164416638900 (Accessed 12 January 2022).

Statista. 2020. Total population of South Africa 2019 by gender. Available: https://www.statista.com/statistics/967928/total-population-of-south-africa-by-gender (Accessed 5 May 2022).

Swanson, R.A., and Chermack, T. J. 2013. *Theory building in applied disciplines*. Montgomery Street: Berrett- Koehler Publishers.

Sweeper, D. and Ramsden, K. 2020. Establishing and promoting an institutional repository and research information management system. *Library Hi Tech News*, 37 (7): 9-12

Tamizhchelvan, M. and Anbalagan, M. 2020. Indian Research Information Network System (IRINS): an analysis of faculty profiles of the Gandhigram Rural Institute – Deeded to be University. *Library Philosophy and Practice*, 1-22.

Tashakkori, A. and Teddlie, C. 1998. *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks: Sage publishers.

Tashakkori, A. and Teddlie, C. 2003. *Handbook of mixed methods in social and behavioural research*. Thousand Oaks: Sage publishers.

Tavakol, M., and Dennick, R. 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education.* 2: 53-55. Available: http://dx.doi.org/10.5116/ijme.4dfb.8dfd (Accessed 5 May 2022).

Taylor, A. 2004. *The Organization of Information*. 2nd ed. Westport, CN: Libraries Unlimited.

Taylor, N., Clay-Williams, R., Hogden, E., et al. 2015. High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Services Research*, 15: 244. Available at https://10.1186/s12913-015-0879-z (Accessed 20 June 2021).

Teddlie, C. and Tashakkori, A. 2009. *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks: Sage Publishers.

Tenopir, C., Sandusky, R. J., Allard, S. and Birch, B. 2012. Academic librarians and research data services: Preparation and attitudes. *IFLA Journal*, 39(1): 70-78.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R.J. and Allard, S., 2017. Research data services in European academic research libraries. *LIBER Quarterly*, 27(1): 23-44. Available: http://doi.org/10.18352/lq.10180 (Accessed 5 May 2022).

Telenko, C., Wood, K., Otto, K., Elara, M. R., Foong, S., Pey, K. L., Tan, U. X., Camburn, B., Moreno, D. and Frey, D. 2016. Designettes: an approach to multidisciplinary engineering design education. *Journal of Mechanical Design,* 138: 02. Available at 10.1115/1.4031638 (Accessed 20 June 2022).

Thomas, J. 2011. Future-proofing: The academic library's role in e-Research support. *Library Management*, 32(1): 37-47. Available: http://dx.doi.org/10.1108/01435121111102566 (Accessed 12 January 2022).

Thompson, B. 2000. Ten commandments of structural equation modeling. In*:* Grimm, L. G. and Yarnold, P. R. (Eds.), *Reading and understanding more multivariate statistics*. Washington, DC: American Psychological Association, 261−283.

Torraco, R. J. 2016. Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4): 404-428. Available: http://dx.doi.org/10.1177/1534484316671606 (Accessed 12 January 2022).

Trochim, W. M. and Donnelly, J. P. 2006. Research methods knowledge base. Available: http://www.anatomyfacts.com/research/researchmethodsknowledgebase.pdf (Accessed 12 January 2022).

Ülikool, T. 2020. Data repositories. Available: https://sisu.ut.ee/rdm_course1/data-repositories (Accessed 12 January 2022).

Tsabedze, V. 2019. A framework for the management of E-records in higher education institutions: a case study of the institute of development management. *Mousaion*: *South African Journal of Information Studies*. 37 (3): 1-28. Available at https://doi.org/10.25159/2663-659X/6461 (Accessed 12 October 2021).

Tu, Y. and Shen, W. H. 2007. Visualizing Changes of Hierarchical Data using Treemaps. In: *IEEE Transactions on Visualization and Computer Graphics*. 13(6,): 1286-1293, Available at https://:10.1109/TVCG.2007.70529 (Accessed 12 September 2021).

UK Data Archive. 2017. Create and Manage Data: Quality Assurance. Available at: http://www.data-archive.ac.uk/create-manage/format/quality (Accessed 12 January 2022)

UK Data Service, (nd). Disposal of data. Available at: https://ukdataservice.ac.uk/learning-hub/research-data-management/store-your-data/disposal/ (Accessed 8 December 2021).

UNESCO. 2021. Report: Recommendations on Open Science. Available: https://www.unesco.org/en/legal-affairs/recommendation-open-science (Accessed 12 January 2022).

United States Congress. 2019. Foundations for Evidence-Based Policymaking Act of 2018. Public Law No: 115–435. Available at https://www.congress.gov/bill/115th-congress/house-bill/4174/text (Accessed 13 January 2022).

United States Department of Education 2014. Privacy Technical Assistance Center. Available: www.dol.gov/sites/dolgov/files/ETA/wioa/pdfs/Privacy_Technical_Assistance_Center-brief.pdf (Accessed 1 August 2022).

University System of Georgia 2015. Learning object repository. Available at https://wiki.ucop.edu/download/attachments/3737/lor05_17_06 (Accessed 24 April 2022).

Uzwyshyn, R. 2016. Online Research Data Repositories: The What, When, Why and How. Computers in Libraries, 36(3): 18-21. Retrieved from https://hdl.handle.net/10877/7597 (Accessed 1 August 2022).

Van den Eynden, V. 2013. Data life cycle and data management planning. Available: https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr_2013.pdf (Accessed 1 August 2022).

Van Wyk, J. and Van der Walt, I. 2014. Going full circle: Research Data Management at University of Pretoria. Presentation at e-Reseach Africa 2014 Conference, held at the University of Cape Town, Cape Town, South Africa. 23-27. November 2014. Available: http://www.eresearch.ac.za/sites/default/files/imagetool/images/140/Going_FullCircle_RDM_UP_VanWyk_VanderWalt_26Nov_2014.pdf (Accessed 1 August 2022).

Vehovar, V. and Manfreda, K. L. 2008. Overview: online surveys. In: Fielding, N., Lee, R.M. and Blank, G. *The Sage handbook of online research methods*. London: Sage Publications, 177-194.

Velden T., Lagoze C. 2009. Communicating chemistry. *Nature Chemistry*, 1(9): 673–678. Available at http://10.1038/nchem.448 (Accessed 17 December 2022).

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, K. J. and Rennison, D. J. 2014. The availability of research data declines rapidly with article age. *Current Biology* 24(1): 94–97. Available at http://:10.1016/j.cub.2013.11.014 (Accessed 12 July 2021).

Viswanathan, V. K. 2012. Cognitive effects of physical models in engineering idea generation. Ph.D. Texas A&M University.

Viswanathan, V. and Linsey, J. 2011 Understanding physical models in design cognition: a triangulation of qualitative and laboratory studies. In: *Frontiers in Education Conference (FIE),* 1-6. Available at 10.1109/FIE.2011.6142848 (Accessed 8 0ctober 2022).

Viswanathan, V. K. and Linsey, J. S. 2012 Build to learn: effective strategies to train tomorrow's designers. In: *The ASEE Annual Conference and Exposition,* San Antonio, Texas. Available at https://peer.asee.org/21031 (Accessed 12 September 2021).

Viswanathan, V. K. and Linsey, J. S. 2013. Role of sunk cost in engineering idea generation: an experimental investigation. *Journal of Mechanical Design*. 135(12): 1-22. Available at https://doi.org/10.1115/1.4025290 (Accessed 12 November 2021).

Wahyuni, D. 2012. The research design maze: understanding paradigms, cases, methods, and methodologies. *Journal of Applied Management Accounting Research*. 10(1): 69- 80.

Wall, M. B., Ulrich, K. T., & Flowers, W. C. 1992. Evaluating prototyping technologies for product design. Research in Engineering Design, 3(3):163–177. https://doi.org/10.1007/BF01580518 (Accessed 1 August 2022)

Wallis, J. C., Rolando, E., and Borgman, C. L. 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8(67332): 1-17. Available: https://doi.org/10.1371/journal.pone.0067332 (Accessed 5 July 2021).

Walters, T., & Skinner, K. (2011). New roles for new times: Digital curation for preservation. Washington, DC: Association of Research Libraries. Available at http://files.eric.ed.gov/fulltext/ED527702.pdf (Accessed 12 November 2021).

Wang, H., Law, K. S., Hackett, R. D., Wang, D., and Chen, Z. X. 2005. Leader-member exchange as a mediator of the relationship between transformational leadership and followers' performance and organizational citizenship behavior. *Academy of Management Journal*, 48(3): 420–432.

Weber, R.P. 1990. *Basic content analysis*. 2nd ed. Thousand Oaks: Sage.

Wei, C., Zhao, J., Ni, J., et al. (2023). What does open peer review bring to scientific articles? Evidence from PLoS journals. Scientometrics, 128(5), 2763–2776. Available at https://doi.org/10.1007/s11192-023-04683-9 (Accessed 19 March 2023).

Whitlock, M.C. 2011. Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, 26(2): 61–65. Available: http://dx.doi.org/10.1016/j.tree.2010.11.006 (Accessed 12 June 2022).

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. *Electronic Library and Information Systems*, 49: 382-407. Available: https://doi.org/10.1108/PROG-02-2015-0017 (Accessed 5 July 2021).

Whyte, A. and Tedds, J. 2011. Making the case for research data management. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available:www.dcc.ac.uk/resources/briefing-papers/making-case-rdm#sthash.Y9tAoGFu.dpuf (Accessed 7 July 2021).

Whyte, A. and Wilson, A. 2010. How to appraise and select research data for curation. DCC how-to guides. Edinburgh: Digital Curation Centre. Available: /resources/how-guides (Accessed 5 July 2021).

Wilkinson, D. and Birmingham, P. 2003. *Using research instruments: a guide for researchers*. London: Routledge Falmer.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E. and Bouwman, J., 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1): 1-9.Available: www.nature.com/articles/sdata201618 (Accessed 12 June 2022)

Winchester, C. L., and Salji, M. 2016. Writing a literature review. *Journal of Clinical Urology*, 9(5): 308-312. Available: http://dx.doi.org/10.1177/2051415816650133 (Accessed 12 June 2022).

Witt, M. 2008. Institutional Repositories and Research Data Curation in a Distributed Environment. *Library Trends*, 57(2): 191-201.

Wittig, U., Rey, M., Weidemann, A., and Müller, W. 2017. Data management and data enrichment for systems biology projects. *Journal of Biotechnology*, *261*: 229-237. Available: http://dx.doi.org/10.1016/j.jbiotec.2017.06.007 (Accessed 12 June 2022).

Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: promoting transparency in open science. Scientometrics, 125(2), 1033–1051. Available at https://doi.org/10.1007/s11192-020-03488-4 (Accessed 9 September 2022).

Woolfrey, L. 2009. Archiving social survey data in Africa: an overview of African microdata curation and the role of survey data archives in data management in Africa. (Doctoral thesis), University of Cape Town: Cape Town, South Africa.

Xu, H. 2016. Faculty use of a learning object repository in higher education. *VINE Journal of Information and Knowledge Management Systems*, 46 (4): 469-478.

Yang, M. C. 2005. A study of prototypes, design activity, and design outcome. *Design Studies* 26: 649–669.

Yin, R.K. 2014. Case study research: design and methods. Thousand Oaks: Sage Publishers.

Yoon, A. 2015. Data reuse and users' trust judgments: Toward trusted data curation (Doctoral dissertation). The University of North Carolina at Chapel Hill): USA. Available at https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83 (Accessed September 14, 2022).

Yoon, A., and Schultz, T. 2017. Research data management services in academic libraries in the US: A Content analysis of libraries' websites. *College & Research Libraries, 78*(7): 920-933. Available: https://doi.org/10.5860/crl.78.7.920 (Accessed 13 March 2022).

Youmans, R. J. 2011. The effects of physical prototyping and group work on the reduction of design fixation. *Design Studies* 32(3): 115-138.

Yunus, M. and Kadir, I. K. A. 2017. Development of a generic model for the preservation of primary research data-based digital resources life cycle. *International Journal of Academic Research in Business and Social Sciences*. 7 (11): 1-24.

Zibani, P., Rajkoomar, M. and Naicker, N. 2021. A systematic review of faculty research repositories at higher education institutions, *Digital Library Perspectives,* 38(2): 237-248. Available: https://doi.org/10.1108/DLP-04-2021-0035 (Accessed 13 March 2022).

Zibani, P., Tomic, N. and Greene, M. 2022. Negotiating Transformative Agreements as an Individual Institution: KAUST Library's Context, Challenges, and Prospects. International Federation of Library Associations and Institutions (IFLA): 87th IFLA World Library and Information Congress (WLIC) Dublin, Ireland, 1-10. Available at https://repository.ifla.org/handle/123456789/1972 (Accessed 23 March 2023).

Zimmermann, D. and Grötzbach, L. 2007. A requirement engineering approach to user-centered design. In: J*acko, J.A. (eds) Human-Computer Interaction. Interaction Design and Usability.* HCI 2007. *Lecture Notes in Computer Science*, 4550. Springer, Berlin, Heidelberg. Available: https://doi.org/10.1007/978-3-540-73105-4_40 (Accessed 13 March 2022).

ANNEXURE A- RESEARCH PROJECT PLAN

Annexure A

Research Project Plan

Research Project Activity	**Timelines**
Preliminary Literature Review	**Jan 2021**
Proposal Preparation	**Jan- Mar 2021**
Literature Review	**Mar-2021-May 2022**
Data Collection	**Feb- Jun- 2022**
Statistical Analysis	**Jul- 2022- Jan-2023**
Journal Publication	**Apr- Jul- 2022**
Writing Thesis	**Nov- 2022- Jan 2023**
Proof reading by 2 academics	**Feb- Mar 2023**
Final editing by supervisor	**Mar- 2023**
Intent to submit	**Feb- 2023**
Submit thesis for examination	**May- 2023**

ANNEXURE B-TURNITIN REPORT

9 April 2023

ORIGINALITY REPORT

14%	12%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	3%
2	researchspace.ukzn.ac.za Internet Source	1%
3	www.emerald.com Internet Source	1%
4	pdfs.semanticscholar.org Internet Source	1%
5	Submitted to Mancosa Student Paper	1%
6	nrl.northumbria.ac.uk Internet Source	<1%
7	www.researchgate.net Internet Source	<1%
8	repository.up.ac.za Internet Source	<1%
9	ir.dut.ac.za Internet Source	<1%

ANNEXURE C-LETTER OF INFORMATION

LETTER OF INFORMATION

Title of the Research Study: The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology

Principal Investigator/s/researcher: Patiswa Zibani, Masters Arts Information Studies

Good Day,

I trust this letter finds you well.

My name is Patiswa Zibani. I am a doctoral student of Library and Information Science in the department of Information & Corporate Management at Durban University of Technology.

I would like to invite you to participate in the study to be conducted on postgraduate researchers from DUT, Faculty of Accounting & Informatics, who have graduated between the years 2015-2020 as well as 32 supervisors who are role players in research administration and management in the faculty. For the accomplishment of doctoral degree the following study must be undertaken:

> "The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology"

Supervisor/s: Dr. Mogiveny Rajkoomar, PhD LIS Co-supervisor: Dr N K Naicker, PhD

Outline of the Procedures

Academic and research institutions are the major role players in the production of scholarly works emanating from their scholars, researchers, and academic staff. However, much focus has previously been more on research publication outputs than research data which informs the outcome of the scholarly work. The purpose of this study is to conduct an audit survey on the management of research data sets at DUT, Faculty of Accounting and Informatics.

The study will employ the use of online questionnaires to postgraduate students graduated between the period of 2015-2020 and interview 32 supervisors attached to the faculty. The outcomes of this study will be beneficial to the university as it will provide an insight on the research data management practices of the Faculty of Accounting & Informatics and advise on the design of a prototype for a research data repository platform which can be adopted by other faculties across the university and ultimately grow data management principles and practices.

The objectives of the research are as follows:

- To identify, classify, and locate research data sets produced by the Faculty of Accounting & Informatics
- To ascertain the format and the management of the existing research data sets in the Faculty of Accounting & Informatics
- Establish the value and maintenance of the research data sets in the Faculty of Accounting & Informatics
- Design a prototype for a research data repository platform for the Faculty of Accounting & Informatics

6 August 2020

A list of postgraduate students' population will be obtained through the faculty office and emails with online survey questionnaires will be sent to them once permission is granted. The list of postgraduate supervisors will also be obtained through the faculty office and appointments will be set up for online interviews requests. The duration to complete postgraduate students online survey questionnaires will be approximately 20 minutes and for postgraduate supervisors, a minimum of 30 minutes will be requested for each interview. The survey questionnaires will be sent out at the beginning of August

2021 and participants will be requested to send back before 31 August. Interviews for postgraduate supervisors will be set up from August to September 2021

Please note the following important information:

a) Participation in the study is completely voluntary, anonymous and participants can exercise their right to withdraw anytime they wish to do so. There are no penalties in any way whatsoever.

b) The study has been reviewed and approved by the Durban University of Technology and will be conducted in accordance with the ethical conditions of the university.

c) You may request that I provide you with more detailed information on all of the ethical conditions and requirements that will be adhered to in this research study.

d) An electronic questionnaire survey method will be used to ensure that the Durban University of Technology, or the researcher, will not be able to identify you or any other individual participant from the information that you will be sharing in this research survey.

e) The information received during the project will only be used for research purposes.
f) You are under no obligation to participate in the study and should not participate in the study if you do not want to. Should you be willing to participate, the informed consent will be requested.

g) The participant will not receive any monetary or other types of remuneration
h) No injuries will be sustained, as the nature of the study does not involve any risk or adverse reaction

The research results will contribute to knowledge the importance of research data management practices at faculty level and how proper organisation can maximize visibility of collections. The researcher will target to publish in academic journals and magazines with an intent to share the current challenges and experiences in creating faculty specific research data repositories. Relevant local and national conferences and symposiums will also be targeted as potential publishers of the findings of this research.

The research data collected will be kept in the repository and access will provided to the researcher and supervisors. The institution framework regarding retention and preservation of research data sets will apply.

If you have any questions or concerns or wish to know more about this study, please contact me, Patiswa Zibani

+966 553734558 or 22173771@dut4life.ac.za, my supervisor Dr Mogiveny Rajkoomar on 031 373 6776 or **mogier@dut.ac.za** or the Institutional Research Ethics Administrator on 031 373 2375. Complaints can be reported to the Director: Research and Postgraduate Support Dr L Linganiso on 031 373 2577 or researchdirector@dut.ac.za.

6 August 2020

CONSENT

Full Title of the Study: The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology

Names of Researcher/s: Patiswa Zibani

Statement of Agreement to Participate in the Research Study:

- I hereby confirm that I have been informed by the researcher, (name of researcher), about the nature, conduct, benefits, and risks of this study - Research Ethics Clearance Number: ___________,
- I have also received, read, and understood the above written information (Participant Letter of Information) regarding the study.
- I am aware that the results of the study, including personal details regarding my sex, age, date of birth, initials and diagnosis will be anonymously processed into a study report.
- In view of the requirements of research, I agree that the data collected during this study can be processed in a computerized system by the researcher.
- I may, at any stage, without prejudice, withdraw my consent and participation in the study.
- I have had sufficient opportunity to ask questions and (of my own free will) declare myself prepared to participate in the study.
- I understand that significant new findings developed during the course of this research which may relate to my participation will be made available to me.

Full Name of Participant Date Time Signature / Right Thumbprint

PZibani

I, ________________ (name of researcher) herewith confirm that the above participant has been fully informed about the nature, conduct and risks of the above study.

Patiswa Zibani 05/09/2021

Full Name of Researcher Date Signature

Full Name of Witness (If applicable) Date Signature

Full Name of Legal Guardian (If applicable) Date Signature

6 August 2020

ANNEXURE D- REQUEST FOR PERMISSION

14 July 2021

Dr Linda Zikhona Linganiso

Research Director

Research and Post Graduate Support

Durban University of Technology

Durban

Request for Permission to Conduct Research

Dear Dr. Linganiso,

My name is Patiswa Zibani, a Ph.D. student in Library and Information Science in the Faculty of Accounting & Informatics, Information & Corporate Management department at the Durban University of Technology. The research I wish to conduct for my Doctoral thesis involves "The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at the Durban University of Technology".

I am hereby seeking your consent to conduct an audit survey on the management of research data sets in the Faculty of Accounting and Informatics at Durban University of Technology by distributing online questionnaires to postgraduate students who graduated between the period of 2015-2020 and interviewing 32 Ph.D. supervisors attached to the faculty.

I have provided you with a copy of my proposal which includes copies of the data collection tools and consent and/ or assent forms to be used in the research process, as well as a copy of the approval letter which I received from the Institutional Research Ethics Committee (IREC).

If you require any further information, please do not hesitate to contact me on +966 553734558 or 22173771@dut4life.ac.za.

Thank you for your time and consideration in this matter.

Yours sincerely,

Patiswa Zibani

Durban University of Technology

ANNEXURE E- ETHICS APPROVAL

Faculty Research Office

Durban University of Technology 12 July 2021

Student: Patiswa Zibani Student Number: 22173771

Degree: Doctor of Philosophy in Library and Information Science Email: 22173771@dut4life.ac.za

Supervisor: Dr M. Rajkoomar Supervisor email: mogier@dut.ac.za

Dear Ms Zibani

ETHICAL APPROVAL: LEVEL 2

I am pleased to inform you that the Faculty Research Ethics Committee (FREC) following feedback from two reviewers, has granted preliminary permission for you to conduct your research ‘The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology’.

When ethics approval is granted:

You are required to present the letter at your research site(s) for permission to gather data. Please also note that your research instruments must be accompanied by the letter of information and the letter of consent for each participant, as per your research proposal.

This ethics clearance is valid from the date of provisional approval on this letter for one year. A student must apply for recertification 3 months before the date of this expiry.

Recertification is required every year until after corrections are made, after examination, and the thesis is submitted to the Faculty Registrar.

A summary of your key research findings must be submitted to the FRC on completion of your studies.

Yours sincerely

Dr Trisha Ramsuraj

FREC Deputy Chair

Faculty of Accounting and Informatics Durban University of Technology Ritson Campus

Durban, South Africa, 4001

ANNEXURE F- GATEKEEPERS LETTER

Directorate for Research and Postgraduate Support

Durban University of
Technology Tromso Annexe,
Steve Biko Campus
P.O. Box 1334, Durban 4000

Tel.: 031-
3732576/7
Fax: 031-
3732946

20th July 2021

Ms Patiswa Zibani

c/o Department of Information and Corporate Management Faculty of Accounting and Informatics Durban University of

Technology Dear Ms Zibani

PERMISSION TO CONDUCT RESEARCH AT THE DUT

Your email correspondence in respect of the above refers. I am pleased to inform you that the Institutional Research and Innovation Committee (IRIC) has granted **Gatekeeper Permission** for you to conduct your research “The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology” at the Durban University of

Technology. **Kindly note that this letter must be issued to the IREC for approval before you commence data collection.**

The DUT may impose any other condition it deems appropriate in the circumstances having regard to nature and extent of access to and use of information requested.

We would be grateful if a summary of your key research findings would be submitted to the IRIC on completion of your studies.

Kindest
regards.

Yours
sincerely

DR LINDA ZIKHONA LINGANISO
DIRECTOR: RESEARCH AND POSTGRADUATE SUPPORT DIRECTORATE

ANNEXURE G- ETHICS CERTIFICATE

TRREE

Clinical Trials Centre
The University of Hong Kong

Zertifikat

Certificat

Certificado

Certificate

Promouvoir les plus hauts standards éthiques dans la protection des participants à la recherche biomédicale
Promoting the highest ethical standards in the protection of biomedical research participants

Certificat de formation - Training Certificate

Ce document atteste que - this document certifies that

Patiswa Zibani

a complété avec succès - has successfully completed

Introduction to Research Ethics

du programme de formation TRREE en évaluation éthique de la recherche
of the TRREE training programme in research ethics evaluation

Release Date: 2021/05/30

CID : gRC1FaSCE7

Professeur Dominique Sprumont
Coordinateur TRREE Coordinator

FMH Continuing Education Program (5 Credits)
Programme de Formation Continue (5 Crédits)

Foederatio Pharmaceutica Helvetiae FPH Programmes de formation continue

Continuing Education Programmes
Programmes de formation continue

Ce programme est soutenu par - This program is supported by :
European and Developing Countries Clinical Trials Partnership (EDCTP) (www.edctp.org) - Swiss National Science Foundation (www.snf.ch) - Canadian Institutes of Health Research (http://www.cihr-irsc.gc.ca/e/2891.html) - Swiss Academy of Medical Science (SAMS/ASSM/SAMW) (www.samw.ch) - Commission for Research Partnerships with Developing Countries (www.kfpe.ch)

[REV : 20170310]

ANNEXURE H- QUESTIONNAIRE

QUESTIONNAIRE GUIDE FOR POSTGRADUATE STUDENTS

Dear Respondent,

I am a student at the Durban University of Technology, currently studying towards the Doctor of Philosophy in Library and Information Science degree.

Title of the study: The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at the Durban University of Technology".

The study seeks to provide an insight into the research data management practices of the Faculty of Accounting & Informatics to inform the design of a prototype for a research data repository platform for the faculty.

Instructions:

- Please respond to the following questions to the best of your ability and as honestly as possible.
- Tick appropriate boxes/brackets and give details where spaces are provided.
- Further instructions are given in some questions.
- This survey will run using any web browser.
- Your cooperation in completing this survey is highly appreciated.
- Confidentiality is assured.

SECTION A: BACKGROUND INFORMATION

1. What is your gender?

Male [ ]

Female [ ]

2. To which Department do you belong?

----

3. What is your highest qualification?

Masters [ ]

PhD [ ]

SECTION B: RESEARCH DATA CLASSIFICATION, MANAGEMENT, MAINTENANCE, AND TECHNICAL ASPECTS

Identification, classification, and location of research data practices

4 What type of research data did you generate for your postgraduate research studies at the DUT Faculty of Accounting & Informatics?

5 Which of the following classification levels did you apply to your research data?

High (Extremely sensitive individually identifiable information) [ ]
Moderate (Moderately sensitive individually identifiable information) [ ]
Low (Non-public, non-sensitive information and de-identified information) [ ]
Minimal (Public information) [ ]

6 Which of the following digital facilities have you used as the location for your research data sets?

Select all that apply**.**

Personal computers [ ]

Office computers [ ]

External hard drives [ ]

CDs for backup. [ ]

Institution's available networked capacity [ ]

Commercial software or services [ ]

Freely available software or services (Google Drive) [ ]

Flash/USB drive [ ]

Email account(s) [ ]

Others (Specify)--

Research data file formats, size, and management practices

7 Which of the following data format(s) was your research data presented?
Please select all that apply.

Digital text or digital copies of text [ ]

Digital images or digital copies of images [ ]

Audio recordings [ ]

Video recordings [ ]

Spreadsheets [ ]

Digital databases (e.g., surveys, census data, government statistics, etc.) [ ]

Computer code [ ]

Biological/organic/inorganic samples or specimens [ ]

Spatial data [ ]

Artistic products [ ]

Others (Specify): --

8 Please estimate the size/volume of research data you have produced through your research project.

1 GB (gigabyte) or less [ ]
More than 1 GB but less than 100 GB [ ]
More than 100 GB but less than 1 TB (terabyte) [ ]
More than 1 TB but less than 100 TB [ ]
More than 100 TB but less than 1 PB (petabyte) [ ]
More than 1PB [ ]
I don't know [ ]

9 How frequently is your research data backed up?

Daily [ ]
Weekly [ ]
Monthly [ ]
Annually [ ]
Ad Hoc [ ]
Don’t know [ ]
Never [ ]

10 If it is backed up, where is it backed up?
Please select all that apply.

Personal computers [ ]
Office computers [ ]
External hard drives [ ]
CDs for backup. [ ]
Institution’s available networked capacity [ ]

Commercial software or services [ ]

Freely available software or services (Google Drive) [ ]

Flash/USB drive [ ]

11 Di you put in place some strategies to protect your research data from loss?

Yes [ ]
No [ ]

11.1 If you answered **No** to question 11, please provide the reasons?

11.2 If you answered **Yes** to question 11, which of the following strategies have you adopted to protect your data from loss?
Please select all that apply.

Copies are uploaded on Goodge Drive [ ]
Copies are uploaded on Drops Box [ ]
Copies are kept in my email [ ]
Copies of data sets are saved on a disk, USB drive, tape, computer hard drive [ ]
Copies of data sets are saved on a local server [ ]
Copies of data sets are saved on a central campus server [ ]
Copies of data sets are saved on a web-based or cloud server [ ]
Copies of data sets are stored in a data repository or archives [ ]
Backup files are automatically generated [ ]
Backup files are manually generated [ ]

Others (Specify): ---

Value and maintenance practices

12 Do you usually share your research data with other researchers or stakeholders?

Yes [ ]
No [ ]

13 If you answered **Yes** to question 9, which of the following factors motivate or compel you to share your research data?
Select all that apply.

Journal policies require me to submit my manuscripts with data [ ]
Research funders compel me to share data from research projects they have funded [ ]
My university requires me to share the data from my research projects [ ]
I share data because open access proponents have advised me to do so [ ]
I share data because I personally find it scientifically necessary [ ]

14 If you answered **Yes** to question 9, which of the following best represent the ways you share your data? **Please, select all that apply.**

Sharing practices	**All**	**Most**	**Some**	**None**
Through external drives (flask disks)				
Through emails				
Through e-journals' websites				
On social networks				
On my personal website/blogs/wikis				
Through clouds (Google Drive, DropBox, etc)				
University repositories				
Through research funders website				
On my faculty's website				
On the principal investigator's website				
Through a national network				
Through a regional network				
Through a global network				
Other (Specify)				

15 For each of the following factors, indicate the extent to which they discourage you from sharing your research data with other researchers.

Statement	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
There is no place to put the data					
Lack of incentives					
Lack of funding					
Lack of standards or guidelines for sharing data					
The data is not fully documented					
License agreements prohibit sharing data					
I would lose control over my data					
I have insufficient skills to make my data available to the public					
The data is in a format that is not widely readable					
My data may be misinterpreted. by others					
The university owns the data I produce					
If funded, the funding agency owns the data					

Insufficient time					

16 For each of the following conditions, indicate the extent to which they could encourage you to share your research data with others.

Conditions	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
I would be willing to place at least some of my data into a central data repository with no restrictions					
I would be willing to place all my data into a central data repository with no restrictions					
I would be more likely to make my data available if I could place conditions on access					
I would be willing to share data across a broad group of researchers who use data in different ways					
It is important that my data are cited when used by other researchers.					
It is appropriate to create new datasets. from shared data					

Others (Specify)					

Research data reuse practices

17 How frequently do you use research data produced/created by other researchers or research institutions in your research activities?

Always	[ ]
Frequently	[ ]
Occasionally	[ ]
Seldom	[ ]
Never	[ ]

18 For each of the following factors, indicate the extent to which they discourage you from using research data produced by other researchers or research institutions.

Factors	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Difficult to find, discover, or access reusable data					
Hard to integrate with my own data					
Not trusting others' collection methods					
Data may be misinterpreted due to complexity of the data					
Lack of common or standard formats					

Lack of adequate metadata/data description information					
Data may be misinterpreted due to poor quality of the data					
Data may be used in other ways than intended					
Legal/ethical restrictions					
Other (specify)					

19 Do you think it is necessary to preserve/maintain your research data?

Yes [ ]

No [ ]

19.1 If you answered **Yes** to question 17, for how long do you think your data will remain valuable?

Indefinitely [ ]

10 – 20 years [ ]

5–10 years [ ]

3–5 years [ ]

1-2 years [ ]

Not sure [ ]

Design research data repository platform prototype practices

20 In your opinion, do you think your faculty provides infrastructure to support your research data management?

Yes [ ]

No [ ]

20.1 If you answered **Yes** to question 18 above, explain the kind of support offered.

20.2 If you answered **No** to question 18, to which extent do each of the following kind of support would you like your faculty to provide?

Kind of support	Agree Strongly	Agree Somewhat	Neutral	Disagree Somewhat	Disagree Strongly
Should establish a process for managing data during the life of the project (short-term – 5 years or less)					
Should establish a process for managing data beyond the life of the project (long-term beyond 5 years).					
Should establish necessary tools and technical support for data management during the life of the project (short-term – 5 years or less)					
Should establish necessary tools and technical support for data management data beyond the life of the project (long-term - beyond 5 years).					

Should establish necessary funds to support data management during the life of a research project (short-term -5 years or less)					

Please, feel free to make any comments in relation to research data management at your faculty.

END OF QUESTIONNAIRE

THANK YOU FOR TAKING PART IN THIS STUDY

ANNEXURE I- INTERVIEW SCHEDULE

INTERVIEW GUIDE FOR POSTGRADUATE RESEARCH SUPERVISORS

Dear Interviewee,

My name is Patiswa Zibani, a student at Durban University of Technology, currently studying towards the Doctor of Philosophy in Library and Information Science degree.

Title of the study: The design of a prototype for a research data repository platform for the Faculty of Accounting and Informatics at Durban University of Technology".

The study seeks to provide an insight on the research data management practices of the Faculty of Accounting & Informatics to inform the design of a prototype for a research data repository platform for the faculty.

Instructions:

• Please respond to the following questions to the best of your ability and as honestly as possible

• Confidentiality is assured.

SECTION A: PERSONAL INFORMATION

1. Faculty: …………………………………………………

2. Department: …………………………………………….

3. Programme……………………………………………...

4. What is your current designation? ………………………

• How long have you been working in the Faculty of Accounting & Informatics?

• How important is research in your professional life?

SECTION B: RESEARCH DATA CLASSIFICATION, MANAGEMENT, MAINTENANCE AND TECHNICAL ASPECTS

Identification, classification, and location of research data practices

• Research can be seen as one of the core activities of a faculty within a university. What are the primary types of research data has been produced by the students that you have supervised over the years?

• How would you describe the current classification process (codes and levels) when it comes to your students research output

• Describe or explain the current facilities or locations that are available at the Faculty for students to store research output.

Research data file formats, size, and management practices

• What are the primary formats that research data sets are presented?

• Would you be able to comment on the file size of these data sets?

• Can you describe the current data management process of the research datasets inclusive of the backup process?

Value and maintenance practices

• Describe your role in promoting strong research data management and practice in your faculty?

• How do you encourage sharing of research data within the faculty/university?

• How does your office view the concept of research data reuse?

• What measures has the faculty put in place to ensure that digital research data is properly preserved for longevity?

Design research data repository platform prototype practices

• In your view, What would be the main factors to consider for when developing a repository structure for research data

• Describe the functionality or usability features that you would envision for a suitable repository for research data

If you have additional comments in relation to the topic under discussion, please feel free to do so.

End of interview.

Please feel free to make any comments in relation to the topic we have discussed.

ANNEXURE J- SAMPLE OF INTERVIEW

Sample of Interview Responses

Question	**Respondent**	**Response**
How important is research in your professional life?	6	Everything we do is confirmed by research; the role of teaching and learning is critical thinking, which is the foundation of research. I cannot entirely call myself an academic if research is not part of what I do and not just a part but a critical component of what I do.
What are the primary types of research data that have been produced by the students that you have supervised over the years	2	Depending on the instruments that the students have adopted for their research studies, it has mainly been a bit of both, that is, qualitative data in the form of interviews, transcriptions, observations and quantitative data in the form of surveys
How would you describe the current classification process (codes and levels) when it	4	In this field, there is a level of restriction to information, but data can be made available upon request from the principal investigator. So in such a case, only the metadata can be provided, not the actual data. The data is then

comes to your students' research data		made available upon request to ensure proper consent. So I would say it depends on the nature of the study.
Can you explain the current facilities or locations that are available at the faculty for students to store research data	5	Supervisors mainly work on the instruments before the student goes to the field to collect data. After the student has collected data, the results sit with the student, not with me as the supervisor or with some central office that makes sure these records are or this data is kept safe and disposed of properly.
What are the primary formats that research data sets are presented	1	It depends on their instruments, but most of my students' data is textual in the social sciences. Looking back at it, my students haven't done much data that will require pictorial information, videos, no, Yeah, so I would say the data primarily is textual.
Would you be able to comment on the file size of these data sets	1	I would not be sure, but as far as I know, I have not had a student who had asked for assistance due to producing extensive data that would require some hostage facility.
Can you describe the current data management process of the research datasets,	1	I have no idea; since it's in the custody of the student and not the faculty or the university, I would not know. In this case, we are at the mercy of students because even as the

inclusive of the backup process	6	supervisor after the students have graduated, we often also hold on to the thesis, that's all. I think there is some level of backup that students do, especially during the research journey, because they would need to reflect on the data now and then. After submitting the research report, I cannot say if there is any backup and maintenance to the research data.
Describe your role in promoting strong research data management and practice in your faculty.	6	Honestly, I have not done much in that area; I have never realized that there can be a role that I can play as a supervisor. I think it's mostly because we have never seen ourselves as role players in research data management until now.
How do you encourage sharing of research data within the faculty/university?	1	I will not say there is any specific activity that I engage in to encourage sharing of data; what I've tended to see is that there isn't a strong inclination by the faculty and maybe in the university to do research from the research data already exists.
How does your office view the concept of research data reuse?	9	It saves time. I can use the same data to present a different take on the study. It also saves on financial resources.

What measures has the faculty put in place to ensure that digital research data is properly preserved for longevity	4	We do not have facilities to keep our research data in the faculty. Normally, students keep their research data.
In your view, what would be the main factors to consider when developing a repository structure for research data?	2	I see this facility being developed in our faculty because the data we could have can help empower other students to avoid redundancy.
Describe the functionality or usability features that you would envision for a suitable repository for research data	9 5	Infrastructure that would accommodate all formats of research data. We need a faculty-driven model as faculties need to understand their role in research data management, and there should be compliance. User perspective should be considered when developing such a repository.

ANNEXURE K- THEMES

Extract of themes definitions

Theme	**Meaning**
Importance of research in professional life	Many of the issues that our society and industry are facing have solutions, and academic research has significantly contributed to that
Data attributes	A data attribute is a quality or feature measured for every observation (a record) and can change from one observation to the next. It could be measured using continuous values (for instance, the amount of time spent on a website) or categorical values (e.g., red, yellow, green).
Data management	The process of gathering, organizing, and gaining access to data to support productivity, efficiency, and decision-making.
Repository development	It involves a collection of objects required to create screens and service components. Reports are defined and stored during development using the visual object repository. You can quickly create a new application component by copying the required objects from the repository once it has been filled.

ANNEXURE L- DELPHI METHODS

Annexure L: Delphi Methods

TITLE: The Design of a Faculty Research Data Repository Platform conducive to a University of Technology
AIM: The aim of the study is to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology
DELPHI method: Round 1 – open-ended questions
Research Objectives
[RO1]: *Analyze the most pertinent studies on research repositories at HEIs.* [RO2]: *Ascertain the management practices of the existing research data in a University of Technology Faculty* [RO3]:

Establish the usefulness of the faculty research data repositories [RO4]: *Design a novel digital prototype for a University of Technology faculty research data repository platform based on research evidence*
1. What current research data management practices necessitate the design of a faculty research data repository? 2. How can user needs be integrated with the design technology of the faculty research data platform model? 3. Identify critical components for the design of the faculty research data repository platform 4. Define design processes necessary for the research data platform architecture of a faculty/ unit.

TITLE: The Design of a Faculty Research Data Repository Platform conducive to a University of Technology
AIM: The aim of the study is to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology
DELPHI method: Round 2 – Questionnaire – statements

Strongly disagree	Somewhat disagree	Neither agrees or disagree	Agree	Strongly agree
1	2	3	4	5

Please mark the statement with a circle that best fits. For each statement, you can only select one option

Number	Statement	Rating				
1.	Literature review analysis provides valuable information on the state of faculty research repositories in HEIs and can be used as an assessment tool	1	2	3	4	5
2.	Understanding research data creators' perspectives is essential for the design phase of the faculty research data repository	1	2	3	4	5
3.	Understanding research data consumers' perspectives is essential for the design phase of the faculty research data repository	1	2	3	4	5

4.	Understanding the research data management practices of the faculty is crucial to inform the repository design	1	2	3	4	5
5.	An insight into the usefulness of a research data repository platform is critical to inform the design	1	2	3	4	5
6.	Critical functional components should be based on user requirements	1	2	3	4	5
7.	Testing a design model with the experts' community for evaluation, critique, and enhancements is essential.	1	2	3	4	5
8.	Testing a model will provide the researcher with an opportunity to either modify the model	1	2	3	4	5

TITLE: The Design of a Faculty Research Data Repository Platform conducive to a University of Technology
AIM: The aim of the study is to investigate research data management in a faculty context in order to support researchers by designing a faculty research data repository platform conducive to a University of Technology
DELPHI method: Round 3 – Revised Questionnaire - statements

Strongly disagree	Somewhat disagree	Neither agrees or disagree	Agree	Strongly agree
1	2	3	4	5

Please mark the statement with a circle that best fits. For each statement, you can only select one option

Number	Statement	Rating				
1.	Literature review analysis provides valuable information on the state of faculty research repositories in HEIs and can be used as an assessment tool	1	2	3	4	5

2.	Understanding research data creators' perspectives is essential for the design phase of the faculty research data repository	1	2	3	4	5
3.	Understanding research data consumers' perspectives is essential for the design phase of the faculty research data repository	1	2	3	4	5
4.	Understanding the research data management practices of the faculty is crucial to inform the repository design	1	2	3	4	5
5.	An insight into the usefulness of a research data repository platform is critical to inform the design	1	2	3	4	5
6	Critical functional components should be based on user requirements	1	2	3	4	5

ANNEXURE M- Language Editing

EDITING LETTER

696 Clare Road
Clare Estate
Durban
4091
26 April 2023

To: Whom it may concern

Editing of D.Phil Dissertation: Patiswa Zibani

The Design of a Faculty Research Data Repository Platform conducive to a University of Technology

This letter serves as confirmation that the aforementioned thesis has been language edited. The requisite grammatical conventions have been met.

Any queries may be directed to the author of this letter.

Regards

MP MATHEWS
Lecturer and Language Editor
Mercimathews4@gmail.com
083 676 4778