Understanding the Research Data Management Practices of Researchers in the Department of Historical Studies at the University of Cape Town

Ingrid Thomson JHNING001

Supervisor: Michelle Kahn

A minor dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Philosophy

2024

Declaration

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: Signed by candidate Date: 11/02/2024

Acknowledgements

I would like to express my gratitude to:

My husband, Deven and daughters, Rebecca and Grace for giving me the support while working on my dissertation; And to my mother Biddy Johnson and siblings, Genevieve Michel, Johnathan Johnson, Belinda Johnson and Samantha Ghoorun for their support during this journey.

Michelle Kahn for her patience, and continuous support as a supervisor – it has been a long haul.

The late Alex D'Angelo for his support and constant encouragement to persevere with the dissertation.

Nikki Crowster, Ujala Satgoor, Susanne Noll, Danielle Abrahamse, Engela Britz, Itumeleng Mpofu, Lise-Marie Hattingh, Maureen Chiware, Jeremiah Pietersen and Vuyokazi Jamieson for support and encouraging words, always enquiring about my progress with "Bruno".

The Department of Historical Studies at UCT who were always supportive and encouraging.

Abstract

The aim of the study was to investigate the current research data management practices in the Department of Historical Studies at the University of Cape Town.

The purpose of the study was an exploration of the research data management practices in the Department particularly in light of the University of Cape Town's research data management policy. The objectives were to understand the research practices of the researchers and whether and how the act of research data management is incorporated into the current research practices and perspectives of historical scholarship within the Department and what support they needed from the University of Cape Town Libraries. There is a dearth of literature covering research data management and historical studies and this study adds to the field of research data management in Humanities of which historical studies is a part. The research design employed was that of a case study. Data was collected through semi-structured interviews of individual history staff, while data collected from students was through focus groups. The data was analysed according to themes which were identified by the questions guiding the study, the literature review, the research data management lifecycle and the history research lifecycle theory that supported the study.

The study findings showed that, while not calling the process data management, researchers were already managing their data as part of managing their research within the history research lifecycle. Findings also showed that there was limited awareness of the official research data management policy at the University of Cape Town, and where there was awareness, the policy was seen as a bureaucratic requirement. The researchers also had limited awareness of the support and services the University of Cape Town Libraries provided.

The study recommends that there should be a domain-specific strategy towards research data management to take the disciplinary differences and culture of history into account, including the tailoring of a data management plan for history.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

DCC	Digital Curation Centre
DMP	Data Management Plan
HDML	History Data Management Life cycle
HRL	History Research Life cycle
NEH	National Endowment of the Humanities (NEH) Office of Digital Humanities
NRF	National Research Foundation
OECD	Organisation for Economic Co-operation and Development
RDM	Research Data Management
RIN	Research Information Network
UCT	University of Cape Town
UKDA	UK Data Archive

Chapter 1: Introduction

1.1 Introduction to the Study

This study arose from this researcher's role as subject librarian for the University of Cape Town's Department of Historical Studies in the Faculty of Humanities and a discussion with a visiting scholar in the Department about their research data management practices. Both raised questions for this researcher around what was understood to be 'data' in respect of historical studies, and if, how and when the act of research data management is incorporated into the current research practices and perspectives of historical scholarship. This question is particularly relevant as the University of Cape Town (UCT) adopted a Research Data Management policy in 2018. This policy (University of Cape Town, 2018) requires that researchers and postgraduate students manage, curate, and possibly publish their research data at the end of a research project. The policy states that research practice related to data management must be consistent, support data sharing and must make research data "discoverable, accessible, reusable and interoperable to specific quality standards" (University of Cape Town, 2018:2).

This introductory chapter lists the research objectives of this study, following a definition of research data and research data management in general, followed by discussion on research data and research data management in Humanities, the discipline in which historical studies falls. Research data and its management in historical studies is defined to further contextualise the study. This is followed by brief remarks about support and services before an overview of the research methodology is provided and study delimitations outlined.

1.2 What is Research Data?

Research data is difficult to define, as much depends on institutional and individual disciplinary practices, requirements, approaches and behaviour, research cultures, attitudes and context (Borgman, 2007:120; Harvey, 2010; Jahnke, Asher & Keralis, 2012; Akers & Doty, 2013:6; Pennock, 2014:3; Weller & Monroe-Gulick, 2014; Sabharwal, 2015:95; Schopfel and Prost, 2016:99; Johnston, 2017).

The National Endowment of the Humanities (NEH) Office of Digital Humanities (2018:1) describes research data as all data created, generated or collected by researchers in the course of their work, as well as third-party data which may have originated with the institution or come from elsewhere. The Organisation for Economic Co-operation and Development (OECD)

Principles and Guidelines for Access to Research Data from Public Funding (2007:13) defines research data as "factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings." Research data could be digital or "electronic" (Weller & Monroe-Gulick, 2014:467) such as text, images, sound, websites or databases, but also includes non-digital or analogue formats such as laboratory notebooks (JISC 2019; Pryor, 2012:) or qualitative or quantitative data such as numeric data files, heritage collections, images, multimedia, code, and so on (Borgman, 2015:19). Although Harvey (2010:45) says that the term 'data' applies to any information in digital form, regardless of the context in which it is created, managed and used. Digital data is collected "often in staggering quantities" (Ray, 2014:1). Transcripts of recordings are considered "research data" (Australian National Data Service 2017).

The definition of data varies from discipline to discipline, as not "all researchers see the materials they work with as 'data'" and "not all agree that those materials should even be considered data" (Mohr et al, 2015: 51).

1.3 Research Data Management

Research Data Management is the process of managing data generated during a research project, "from the start of the research cycle through to the dissemination and archiving of valuable results" (Whyte & Tedds, 2011:2). It involves maintaining, preserving and adding value to the research data for the long term which enables access to, and sharing of that data. Andrikopolou, Rowley and Walton (2021:1) describe the "governance of research data through the entire research lifecycle" as Research Data Management.

There are two outputs of Research Data Management: "the long term preservation of datasets through archiving, and the sharing and reuse of datasets for further research and other purposes in society at large" (Johnsson: 2015:5). Management of research data ensures that data can be "discovered and validated" should the need arise (Ray, 2014:7). As Whyte and Tedds (2011:2) note, research data management includes curation and preservation of the data "ensuring project results [data] are fit to archive" and "remain fit for reuse" while remaining "fit for secondary use in the longer term" (Whyte & Tedds, 2011:2).

Data is central to every stage of the research lifecycle, from the beginning stages with planning

of the project, to the data acquisition and analysis to the publication of the research results (Briney, 2015:10). Additionally, data sharing, data preservation and data reuse can be mapped to the research lifecycle, demonstrating that data is an "integral part of research" (Briney, 2015: 11). Briney (2015:3) reminds that "data is a valuable research product that should be treated with care" and "data management requirements are one way to make this happen".

A research data management plan outlines how data will be collected and generated, and how they will be managed throughout, and after the research (Thoegersen, 2018:492). A research data management plan is good research practice ensuring researchers consider data management at all stages of their research, but it is also "intended to be used to ensure compliance at an institutional as well as faculty, school and individual researcher levels" (Shelly & Jackson, 2018: 399).

A distinction can be made between curation and research data management, where data management takes place in the active phase of the data lifecycle, when "researchers are generating and making use of data themselves for their own purposes" while curation refers to all the activities in the data life cycle (Carlson, 2014; Dreswell, 2017). Andrikopolou, Rowley and Watson (2021:4) note that RDM processes are generally "described as digital curation" because most "RDM involves digital data". Higgins (2012:37) points out the aim of digital curation is to enable access, use and re-use of data.

1.4 Research Data and Research Data Management in the Humanities

The National Endowment of the Humanities (NEH) describes Humanities as the study and interpretation of several academic disciplines including language (modern and classical), linguistics, literature, history, jurisprudence, philosophy, archaeology, religion, ethics, arts, and those aspects of social sciences which have humanistic content or use humanistic methods (NEH, nd.). Humanities scholars are often seen as working on their own, focussing as "depth" rather than "breadth" researchers. However, a report from Research Information Network (RIN) Bulger et al found the image of the "lone scholar" was incorrect (Bulger et al, 2011: 0) with humanities researchers working with new tools and technologies collaboratively and producing and using resources in various new ways (Bulger et al, 2011:6).

There has been a major change in humanities research over the last 15 years as a result of possibilities opened by new technologies and the growth of more formal and systematic collaboration in response to funding opportunities (Bulger et al, 2011: 7). Humanities scholars use a wide range of analogue and digital resources for research, from different times and

different types including traditional (printed and manuscripts) to products of new digital research such as digital text corpora or archival records, or artefacts (Bulger et al, 2011:10; Flanders and Muñoz, 2012).

Ray (2014:19) notes that the question of humanities research data has been "often overlooked" in research data management as the focus has been on scientific data. Schoch (2013) and Posner (2015) say often humanists do not consider the resources they work with as data, but as Pryor (2012:2) observes, we all "do data whether we work in science, social science or humanities." Borgman (2007:213) notes humanities are "more interpretative than data driven." The range of research resources described in the preceding paragraph points to the existence of different types of humanities research data. This makes the characterisation of humanities data difficult.

Data in the humanities are described as "particularly ambiguous" (Borgman 2015:27) while Schoch (2013:3) observes "data in the Humanities is a bit special" and is problematic in that objects of study are not regarded as "data" but rather as books, films, texts, images and sounds, but are essentially 'data' generated or collected during research. Bulger et al. (2011:5) point out researchers in art and humanities do not publish a "great deal of research data" preferring to publish in monographs, but there is a "growing number who produce and share their datasets".

As Owens (2011, paragraph 2) observes, data for a humanist can be described as a "multifaceted object which can be mobilised as evidence in support of an argument." He defines data for humanities as "an artifact or a text that can hold the same potential evidentiary value as any other kind of artifact" (Owens, paragraph 10), noting "the kinds of questions humanists ask about texts and artifacts are just as relevant to ask of data" (Owens, 2011, paragraph 11). Researchers ought to be able to "discover" data through searching and browsing

Borgman (2015:27) notes primary and secondary sources in humanities research are the closest "to raw and processed" data in the sciences and social sciences. Salo (2020:215) describes 'research data' as evidence noting that many humanities researchers "find that phrase [research data] unintelligible with respect to their own work". Di Cresce and King (2017:227) ask the question about how data coming out of "humanistic inquiry" is managed when it is not as "mathematically measurable and regular as scientific data." Indeed, Schoch (2013:3) observes that "data" in the humanities has been "challenged" by mainstream scholars who see data-driven research at odds with "humanistic inquiry." He has developed a definition of data in the humanities which responds to this challenge: "a digital, selectively constructed, machine-

actionable abstraction representing some aspects of a given object of humanistic inquiry" (Schoch, 2013:3). Willaert and Cottyn (2019:15) observe that within humanities, there is a perception with some scholars that "data" is synonymous with "digital data" and thus is associated with the "hard sciences".

Ruediger and MacDougall (2023:8) note that humanities scholars consider their research as being "based on sources rather than data", Moulaison-Sandy and Wenzel (2023:70) add that "humanities might look askance at the idea of having data" even if they use data and if the data they collect is seen by others as data.

To add to the layers of complexity around research data management, Akers and Doty (2013: 6) point out that not only are the needs different across disciplines, but research cultures including practices and different attitudes create complex layers. Humanities research methodologies contain curatorial practices in which data and resources are managed, but these practices are not always recognised as research data management practices (Flanders & Munoz, 2012). It seems much depends on the individual research practices within a humanities discipline or where "the adoption of new technologies is an extension of existing research practice" (Bulger et al, 2011:13). Willaert (2019:9) concurs, referring to the "many unwritten rules, best practices and conventions" in the different humanities disciplines or "even individual departments". These point to how humanities researchers "locate, evaluate, organise, manage, transform and communicate their research data (and information sources) as part of the research process" (Bulger et al, 2011:4). JISC (2019) notes that managing research data is "probably already part of the research process" followed by the researcher (JISC, 2019).

According to Ruediger and MacDougall (2023:3), except for digital humanities researchers, fewer than 20% of humanities researchers report depositing and sharing research or datasets as part of their research practice.

1.5 Research Data and its Management in History

Jordanova (2019:35) describes 'historical research' as the "discovery of fresh materials and the articulation of novel insights" by historians applying their "thinking and puzzling" skills, then applying writing skills. The historian researches the available evidence to "construct a story" (Williams, 2020:8) "based on sources and evidence that support that argument" (Williams, 2020:47). Crymble (2021:18) defines historical research as an "argument-driven, evidence-based answer to a question about the past". Edmond (2016:99) notes that historians "pride

themselves on being ‘source-led’ and characterise the process by which they define research questions as one of finding a ‘gap’ in the current research landscape”.

Historical scholarship, which falls within the broad discipline of the humanities, is divided into a number of subfields or branches related to criteria such as time periods; research methods such as oral history; places and people, for example, women’s history or family history, and other topics (Williams, 2020; Jordanova, 2019).

The historian works with the available evidence which may “include written records, archives, manuscripts, maps, documents” and unwritten evidence such as photographs, paintings, records, tapes, videos, computer hard drives and so on (Williams, 2020:9). They work with primary and secondary sources termed “crucial tools of the historian” (Williams, 2020:47).

Sources and the evidence of these are “the raw material of history and the historian’s most valuable tool” (Williams, 2020:47). Primary sources are usually archival while secondary sources relate to publications or works by other historians in journal or monograph form.

Both these sources represent the ‘data’ that historians use. These sources are found, largely, in archives and libraries, with the ongoing digitisation of primary sources and the presence of born-digital documents, making it easier for historians to access research material so “historical scholarship increasingly depends on our interactions with data” (Gibbs & Owens, 2012). Williams (2020:67) commented that “historians depend on the primary and secondary work of other historians to do their work” and that the work should be transparent, and repeatable or open to interpretation by other historians which means that the data or sources should be accessible or available, for which research data management activities should be practiced.

Jordanova (2019:36) notes there are “difficulties” accessing primary sources, notwithstanding the ongoing digitization, as not all collections have been “digitised, microfilmed, or transcribed” or even catalogued, existing in print form only, while some archives or institutions are not easily accessible physically. For historians, as Jordanova (2019:42) notes, work with sources is “a fundamental part of historical practice” and these sources, as described above, are their data. Edmond (2016:94) describes the historian’s data as “source material bearing evidence and witness to events of the past, objects that stand at the beginning of the individual historian’s process of knowledge creation, but which have already been curated and indeed created by other individuals”.

Recent literature speaks about 'digital history' as "an approach to examining and representing the past that works" with computational tools and digital methods, where historians make extensive use of digital techniques and resources (Jordanava, 2019:221), or as Crymble (2021:14) describes it as "doing history in the digital age". Digital history is linked to digital humanities, which uses computational tools and digital methods to answer research questions, which means that these tools and specialised software can be used to query larger datasets (Ter Braak et al, 2016:22; Graham & Milligan, 2016:3). Crymble (2021:75) defines these large or complex datasets as "big data". Examples of big data include United Kingdom (UK) Parliamentary Papers Online and the Old Bailey Proceedings Online. It is important to note that the underlying research methods whether for digital or analogue historical studies remain recognisable even though new tools and digital technologies have been introduced (Rutner and Schonfeld, 2012:1)

Sinn and Soares (2014:1803) comment that the "authenticity of historical research could be proved … if historians provided an easy method for others to check their sources", implying that RDM within historical research practices (regardless of whether it is digital history or analogue historical studies) is essential.

1.6 Support Services for Research Data Management

As Flores et al (2015) and Pinfield, Cox & Smith (2014) note, the importance of RDM is being recognised by various stakeholders including IT services, the research office and libraries to manage and support researchers from all disciplines in their data management across the research lifecycles.

There are a number of roles, opportunities and activities for libraries to provide support for RDM. Roles include providing consultation and technical advice and support related to "training, development and implementation of RDM plans" including creating metadata and uploading data to repositories. (Shelly & Jackson, 2018:406).

1.7 Problem Statement

With the background described above, and with the researcher's professional interest in the Department of Historical Studies which resides in the Humanities Faculty at UCT, the questions arose about whether and how the act of research data management is incorporated into the current research practices and historical scholarship of researchers in the Department of Historical Studies at the University of Cape Town, and what RDM support they need from the

University of Cape Town Libraries. There was no evidence available regarding these questions. This study's objective was therefore to fill that research gap.

1.8 Research Questions

The following research questions have been formulated to respond to the study objectives.

- RQ 1: What are the research practices of the researchers in the Department of Historical Studies at the University of Cape Town?
- RQ 2: To what extent do research data management practices form part of research practices in the Department of Historical Studies at the University of Cape Town?
- RQ 3: What support for research data management is needed by the researchers in the Department of Historical Studies from the University of Cape Town Libraries

1.9 Overview of Research Methodology

The research study uses a qualitative case study approach. The target population for the research were the academics and masters and doctoral students in the Department of Historical Studies at UCT. Data were collected through semi-structured interviews and focus group discussions.

1.10 Delimitations of the Study

Delimitations are aspects of a study within the researcher's control (McGregor, 2019: 38). This study was delimited to the discipline of History. A further delimitation was to the Department of Historical Studies at the University of Cape Town.

1.11 Structure of the Research Report

This study is organised into five chapters. Chapter 1 presents the introduction and background to the study, the research problem, and study objectives and research questions. It includes a brief overview of the study methodology. Chapter 2 covers a review of literature relating to the study. Chapter 3 reports on the research design, methods and processes that were used for carrying out the study. Chapter 4 presents the analysis of the data collected. Chapter 5 discusses the main findings in relation to objectives of the study, the literature review and findings, concluding with recommendations. A reference list of all sources consulted, and relevant appendices form the final part of the report.

1.12 Summary

This chapter introduced the study, presenting the context and background to the research problem. The objectives are presented together with an overview of the research methodology used, and delimitations of the study, as well as the structure of the research report. The next chapter presents a literature review related to the study and outlines the framework that informs the study.

Chapter 2: Literature review

2.1 Introduction

A literature review is the process of finding and assessing literature that relates to the topic, sharing results of other closely related studies, as well as relating the study to the "larger, ongoing dialogue in the literature, filling the gaps and extending prior studies" (Creswell, 2014:28), providing a sound reason for why a particular study should be conducted.

For this study, the literature is used to illustrate where and how research data management fits into current research lifecycle practices within historical scholarship, and what support and services are needed by historians to manage their research data effectively. The literature was sourced from monographs, chapters in books, journal articles and websites including weblogs emanating from North America (USA and Canada), United Kingdom and European Union. Despite the researcher continually seeking the latest literature relating to research data management in historical studies, there was very little literature available on the topic and none from South Africa or Africa.

The chapter has been arranged in four sections that follow the introduction. The first section describes the research lifecycle, segueing into the research lifecycle in history, with the next section exploring research data in historical studies. These sections are in response to research questions 1 and 2 which ask what the research practices of the historians in the Department of Historical Studies are and to what extent RDM practices are part of those research practices. The third section identifies several research data management lifecycles of interest to this study to further understand RDM practices. These lifecycles can inform RDM support and services, specifically those needed by researchers in historical studies, which the last section explores (responding to research question 3 which asks what support for RDM is needed by the Department of Historical Studies).

2.2 Research Lifecycle and the History Research Lifecycle

In this section, the research lifecycle is briefly described. As the research process varies among disciplines and research domains, a discussion on the History Research Lifecycle in particular follows, linking to research practices and activities but highlighting data management practices.

2.2.1 Research Lifecycle

The research lifecycle is the process that a researcher follows when conducting research,

completing a study from start to completion. Cox and Verbaan (2018:16) describe a research lifecycle very broadly as research moving "from ideation, to funding, to permission, to data collection, to data analysis and then to write-up and further dissemination" which may lead to another project that takes ideas further and starts the lifecycle once more. A simple model of the research process, as described by Punch (2000:39) is that the research is framed in terms of the research question, determining what data is needed to answer the questions, followed by the research design to collect, and analyse that data, and using the data to answer the questions.

2.2.2 History Research Lifecycle

In this section, the research lifecycle stages or phases in history are described to place data management in history research in context. Uwa (1977) uses the term 'stages' while the University of London (nd) uses 'phases' to describe the activities along the path of the history research lifecycle.

Figure 1 below (University of London, nd) shows the history research lifecycle outlining the phases from conceptualisation of the project to completion of the research.

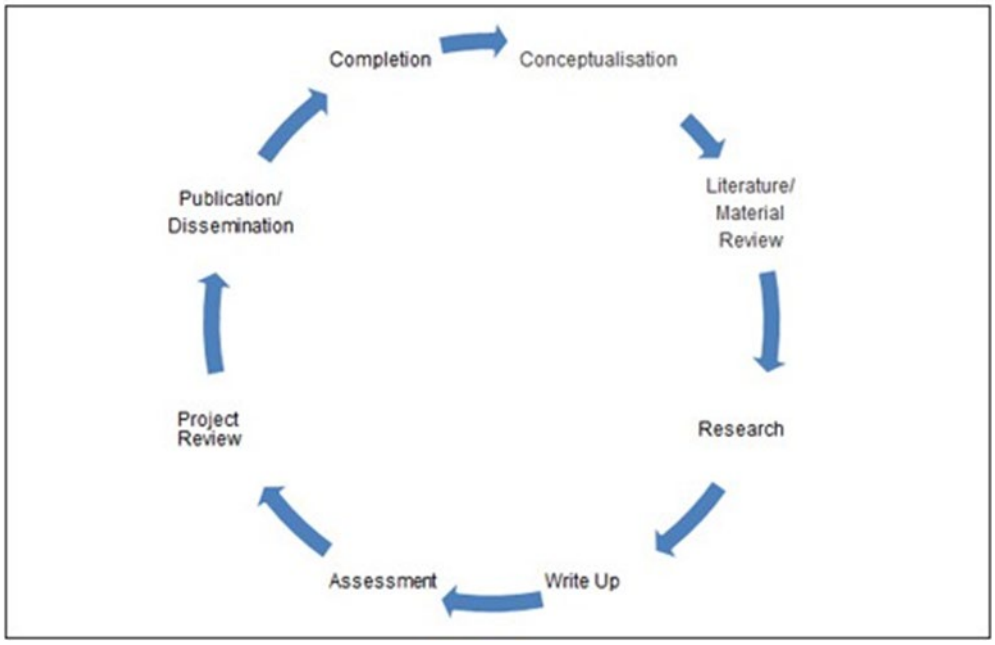

Figure 1: The History Research Lifecycle (University of London, nd)

There are eight phases in the history research lifecycle as described by the University of London (n.d). These phases are a continuous process. The first phase begins with the conceptualisation of the research and is followed by a review of the literature or material. During the literature/material review, preparation is made to carry out the research, including gaining

access to the relevant resources, which is then followed by the research work. Once the research is completed, the 'Write Up' phase follows. This is followed by the assessment by the researcher, or supervisors, or team leaders and other colleagues where checks are made on the research project, which could include constructive criticisms or identification of further points of development. Following corrections or further development, the research project is then reviewed again by the researcher and colleagues. This is then followed by the results of the research to be published or disseminated. The research lifecycle is then complete.

Antonova (2020:41) describes the history research lifecycle as starting with secondary sources to gather "basic facts, evidence and existing interpretations" to identify research questions. This is then followed by locating primary sources to answer these questions, using "specialised databases and reference works to find out what is available and where it is located", to research and then start writing. There is feedback and "brainstorming with colleagues" during this research and writing up, before publication or dissemination of the research (Antonova, 2020: 41).

Uwa (1977:21) identified the five typical stages of the historical research lifecycle as problem selection, detailed planning, data collection, analysis and interpretation of the data and writing and re-rewriting or presentation of the findings. He described the data collection stage for historians as "simply going to the primary sources" (1977:22) which links to the 'Literature Material/Review' and 'Research' phases in the history research lifecycle as described by the University of Hull.

Although the aforementioned recent descriptions and the diagram of the history research lifecycle implies that the lifecycle is linear, it often is multi-directional, as Uwa (1977:22) notes that a historian could be working on several stages of research at the same time.

2.3 Research Practices in Historical Studies

This section of the literature review looks at research practices in historical studies, expanding on the history research lifecycle and includes the activities that are part and parcel of the lifecycle. This provides context for the later sections on data in historical studies, and research data management practices in history.

Research practices consist of "a number of activities, from selecting, evaluating and interpreting historical evidence, to communicating these findings" (McDowell, 2002:11). Several studies

(Uwa, 1977; Trinkle, 1998; Dalton & Charnigo, 2004; Bulger et al, 2011; Rutner & Schonfeld, 2012; Chassanoff, 2013; Hitchcock, 2013; Sinn & Soares, 2014; Graham, Milligan & Weingart, 2016; Milligan, 2019; Ma & Xiao, 2020; Antonova, 2020; Corfield & Hitchcock, 2022) have examined historians' research practices, research methods and use of information sources, and in the later studies, have focussed on digital history.

As Williams (2020:xi) observed, history is not just a search for evidence about the past, but can be ongoing research, with potentially multiple examinations of the evidence, "whether through discovery of new facts, or analysis of new research data" as noted by McDowell (2002:12). Williams (2020:9) noted that the historian "investigates what happened in the past by researching the available evidence" with McDowell (2002:4) commenting that "historians examine the past so that we may have a better understanding of the content of past events and the context in which they took place". Williams (2020:9) spoke about "discovering evidence, but constructing a history", and noted that this evidence included "written records, archives, manuscripts, maps and documents, but also unwritten evidence" such as photographs, painting, tapes, videos, computer hard drives and so on. McDowell (2002:11) noted the selection and arrangement of historical data can produce different conclusions of the same evidence. As Corfield and Hitchcock (2022:13) stated that documenting the evidence is a basic requirement of historical research.

As described in the section on the History Research Lifecycle (Section 2.1.1) the process of research that is followed is the identification of a research question, which comes about after reading the secondary scholarly sources (Antonova, 2020:41). Available or "feasible to be acquired" evidence in the form of primary sources is located, accessed, read, and analysed and written up (Antonova, 2020:41). Antonova further noted "we do not, however, simply write up results, because our reading of the messy, equivocal, contradictory, and fragmentary evidence that is left to us is never so clear" which implies that the evidence collected needs to be managed and gathered (Antonova, 2020:41) and properly documented (McDowell, 2004:5).

Historical research is conducted in archives as the main source for primary source materials while libraries are used for secondary source materials (Rutner & Schonfeld, 2012:7). Rutner & Schonfeld note that extensive work is done in the archives using physical and digitised collections (Rutner & Schonfeld, 2012:8). These archival sources are scattered throughout different organisations and in different places (McDowell, 2004:55) and not always described in detail in finding records in archives as an example (Antonova, 2020:41). So, research trips

to archives are a key aspect of research practices for historians (Rutner & Schonfeld, 2012:8).

In a white paper "Digital History and Argument" (Arguing with Digital History Working Group, 2017), it was noted all historians practised the "art of historical interpretation" (2017:3). In a nutshell, all historians select sources and explore those sources to answer their research question, and identify those that provide relevant evidence (ADHWG, 2017:3). The white paper notes further that historians generally are oriented toward print publications but will draw on digital material in their key archives or collections, but in citation practices "continue to favour references to a print source and not the digitised version actually consulted" (ADHWG, 2017: 7).

The ongoing digitisation of primary sources and the growth in born-digital documents makes it easier for historians to access research material so "historical scholarship increasingly depends on our interactions with data" (Gibbs & Owens, 2012). Notwithstanding, a comparatively small number of primary resources have been digitised, so physical visits to archives are essential (Rutner & Schonfeld, 2012:4).

In the study by Ma and Xiao (2020:4) where practices in digital history were examined, one finding was that there a mixed use of conventional and emerging digital methods to collect data, including traditional archival research, digitisation, and crowdsourcing (2020:8).

The results of historical research may be published in journal articles, or chapters in edited books or scholarly monographs (MacDowell, 2002:170). Although he did not refer to historical studies specifically, Williams et al (2009:3) reported that the monograph is the "single most valued means of scholarly publishing and communication" in Arts and Humanities, representing the "gold standard" ahead of journal articles. This notion is echoed by Uva (1977:22). Antonova (2020:42) noted journal articles or essays are seen as smaller projects, while larger projects are published in book form.

2.4 Data in Historical Studies Research

This section of the literature review unpacks the term 'data' in historical studies and historians' attitudes towards 'data'. The term 'data' does not appear regularly in the earlier literature, but rather in the later studies about digital history, which is reflective of the current emphasis on research data.

McDowell's guide to historical research (2002:11) referred to research data, pointing out the "selection and arrangement of historical data" could lead to different conclusions from the same evidence.

Sinn & Soares (2014:1804) report that some of the respondents in their study did not think that primary sources counted as data for a historical study but rather the process of understanding and interpretation of the same was the historical research process itself.

Although Williams (2020:67) does not refer specifically to 'data' in his student guide, he points out that historians depend on the work of other historians to do their work, so work – in other words, the sources, or the data - should "be transparent to, and repeatable by other historians" implying that these are 'data'.

In the Postgraduate Online Research Training Course for History students at the University of London (University of London, nd), it is noted the most common types of data produced by historians is text which could be in the form of notes. These might simply be notes derived from other texts, lectures or jotted down thoughts and ideas, but could also be more structured. Transcriptions and translations are common, as are rough drafts or annotated texts. Other types of data include quantitative, pictorial, and oral data (University of London, nd). Green and Nicholls (2011:4) reported that four types of data were identified by historians namely, qualitative data, quantitative data, oral testimony, and archival data.

The Arguing With Digital History Working Group (2017:13) refers to computational digital history and notes that, once historians begin to consider texts as data, primary sources would be treated as elements of data to be transformed and simplified to answer questions about the past, with the creation of datasets.

Ma and Xiao (2020:6) reported that the historians in their study had difficulties in defining 'data' in a digital history context with a respondent commenting that the term 'data' is not used to describe materials such as videos in a digital collection, in this case, referring to the videos as "video archives".

2.5 Research Data Management Practices in Humanities and Historical Studies

Leading from the previous section which reported on data in Historical Studies, this section describes the RDM practices, firstly in the broader humanities, then in Historical Studies.

2.5.1 Research Data Management Practices in Humanities

As humanities research has become more digital and 'data-driven', especially with the growth of digital humanities, funding organisations such as the National Endowment of the Humanities (NEH) Office of Digital Humanities are asking for data management plans for grant applications, with one of the reasons given to "consolidate information about data to a consistent format" and to "encourage thinking in broader terms about humanities data" (Dressel, 2017: e115:3).

Di Cresce and King (2017:228) observe that the theoretical underpinnings of humanities research cannot be ignored when applying scientific understanding and practices to the field of data management in humanities. This observation does point to the need for sensitivity to practices and cultures of different research communities by funders, universities and service providers' policies and strategies (Bulger et al: 2011:7). However, research methodologies also determine how researchers interact with their data, so even where the theoretical foundation and subject may be different, data practices are likely to be similar, for example, for those doing archival research (Weller and Monroe-Gulick, 2014:470).

In several studies (DCC, 2010; Bulger et al, 2011; Fear, 2011; Janke, Asher & Keralis, 2012; Scaramozzino, Ramirez & McGaughey, 2012; O'Reilly, Johnson & Sanborn, 2012; Akers & Doty, 2013; Weller & Monroe-Gulick, 2014; Awerkamp, Gu and Rogers, 2014; Schopfel and Prost, 2016; Di Cresce and King, 2017; Yu, Deuble, Morgan, 2017; Johnston et al, 2018; Thoegersen, 2018; Wiley and Kerby, 2018; Willaert et al, 2019; Ruediger & MacDougall, 2023), researchers from a range of disciplines including Humanities and Social Sciences were surveyed about data management practices, needs and attitudes and it was observed that these attitudes, needs and practices were closely linked to the discipline in which they work.

Attitudes are affected by structural limitations on sharing data but also by the cultural heritage of individual disciplines (DCC, 2010:3). This idea is echoed by Cox and Pinfield (2014:301) who refer to the differences in academic disciplines and the "corresponding differences in information and data practices" including aspects such as what good data management constitutes. Pinfield, Cox and Smith (2014:25) note that RDM lifecycle models need to adapt to different disciplinary cultures of research communities, observing that, while some disciplines have a "deeply ingrained culture" of sharing or reusing data, or have good data management practices, others such as the arts and humanities, "rarely even use the term 'data'."

These observations indicate the continuum of differences in attitudes and practices towards RDM.

Akers and Doty (2013:9) reported humanities researchers were likely to be unfamiliar with funding agency requirements for a data management plan and were often unaware of the amount of data they were storing and whether these materials were indeed considered data. They also reported humanities researchers were less willing to share their data, whether via a data repository or database, with researchers, expressing concern that they would not be credited for their data (Akers & Doty, 2013: 9).

Bulger et al (2011) noted there was a limited uptake of tools for data management and sharing, with the researchers managing and storing information on their devices and sharing with others via email.

Wolski and Richardson (2015:2) observed that data management practices need to cover the whole research cycle, not just the publishing phase, from "grant inception, to data capture/creation through to archiving" with the challenge that all the data may need to be preserved for future re-use. Salo (2020:217) concurred, observing that, unless there is a "significant preservation effort", data could disappear, but she further observed that preservation is an organisational, local infrastructure and funding issue rather than a technological issue.

Willaert et al (2019:15) noted that one of "key impediments" or "resistance" to RDM is when or where the which frameworks and methods are not seen as possible extensions of the research methods and practices in a particular discipline or field. Furthermore, the requirements of research data management may be seen as additional administrative "burdens" for humanities researchers (Willaert et al, 2019:15).

Moulaison-Sandy and Wenzel (2023:77) commented that humanities researchers (which include historians) require long-term management of their research data in order to use their research as it develops over the years.

2.5.2 Research Data Management Practices in History

What is apparent from the literature is that, as Green and Nicholls (2011:7) report, data management is regarded as "external to the practical aspects of the historian's work" and is seen as "peripheral and optional" rather than as part of a history research project (Green &

Nicholls, 2011:7). In their study, Rutner and Schonfeld (2012:7) do not refer to research data management, instead describing organising and managing research notes as a way for historians to "gain intellectual control over their research topics". Historians working on monograph projects arranged their material into "to-do lists, sources, notes, digital images and other inputs" according to Rutner & Schonfeld (2012:24). These practices echo Salo (2020: 215) describing humanities "research data" as "evidence". Antonova (2020:43) commented that "we note where we found our evidence" so that other scholars can locate them.

Cornfield and Hitchcock (2022:57) noted that history projects are exercises "in controlling a flood of notes and documents" and advised that it is "vital to keep careful notes of what has been examined" and the methodologies used (2022:39). They added that, at an early stage in their research, researchers need to decide on "their preferred strategy for data management" (Cornfield & Hitchcock: 2022:40).

Sinn and Soares (2014:1756) commented that digital technology and resources have had a "significant impact" on the research processes and practices of historians, especially around the use of archival resources, enabling materials to be taken home for later review, which reflects observations made by Rutner & Schonfeld (2012:8).

Researchers who do historical analysis are "more likely to print and save physical copies of text" or keep copies of notes either physically or digitally (Rutner & Schonfeld, 2012; Sinn & Soares, 2014:1796; Weller & Monroe-Gulick, 2014:474). This evidence points to the practice of preserving research data and confirms that the term "research notes management" is an apt description (Rutner & Schonfeld, 2012). Jones (1998:83) describes how it was necessary to manage the data he was collecting for his research by using a computer to capture and organise it, instead of manually capturing on physical index cards. In a similar vein, Corfield and Hitchcock (2022: 8) pointed out that historians have to organise information collected in archives and from secondary readings, and the method used was often an individual choice, but also dependent on the kind of sources being consulted. They noted that, with many more primary and secondary resources available online, it is "much easier to accumulate ever-larger bodies of evidence" (Corfield & Hitchcock, 2022:59) but this leads to difficulties or challenges of managing and locating the data, so it was necessary to have a data management plan (2022:59).

The study by Di Crece and King (2017) explored the data management practices of researchers

in Medieval Studies looking at practices where researchers visited archives. Four key points were identified which confirm the practices mentioned in the preceding paragraph. Scholars had very limited time with which to work with physical manuscripts, especially where items had not been digitised, so that there was very little opportunity to return to the archive if information was missed. The second issue related to the preparations prior to the visit to the archive. The third point was that the researchers tended to produce a "multitude of both digital and analog files" with photographs taken in the archive, a particular challenge (Di Cresce & King, 2017:233). The last point was that researchers develop their own personalised approach or "idiosyncratic data management system" (2017: 234). This point is supported by Wolski and Richardson (2015:2) who observed that researchers may already be managing their data using available tools and technologies, making this a "highly personalised research process for historians" (Rutner & Schonfeld, 2012:23). Corfield and Hitchcock (2022:59) note that, whatever process or data management system is chosen, it is vital that it records all the information needed. Moulaison-Sandy and Wenzel (2023:75) concur, noting that the "personal information" created and under the researcher's control, forms part of the data collected and should be managed.

Ma and Xiao (2020:10) noted that, within digital history, there is a "lack of established preservation plans and awareness", and observed that, compared with "traditional analog historical research", digital history research needs to include the data generated during the "iterative digital processes" and not just the" final results data". The White Paper (Arguing with Digital History Working Group, 2017:19) points to challenges, and increased burdens for the digital historians in that the data, including the original datasets and "other artifacts of their research" has to be shared as a requirement of peer review. In contrast, analog historians only "produce research notebooks in very rare cases of accused plagiarism" (Arguing with Digital History Working Group, 2017:19).

A number of authors (Sinn & Soares, 2014; Arguing with Digital History Working Group, 2017; Ma & Xiao, 2020) observed different attitudes of historians towards data, where, for example, some respondents in those studies did not think primary sources counted as data, but rather the process of understanding and interpreting the data was the historical research process itself. Ma and Xiao (2020) noted different understandings of the term "data" in digital history research based on the relationship that data has with sources, collections and evidence.

Corfield and Hitchcock (2022:60) observed that researchers were likely to have many digital

files and that a system of file naming was essential, whether it was for photographs, websites or documents. This practice is one of the ways in which researchers are able to manage their data. As they noted

> *good data management allows researchers to organize their thoughts as they proceed, as well as to organize their writing; and, above all, to think more critically about what the project is actually trying to achieve. With the right tools, researching and writing becomes infinitely more efficient* (Corfield & Hitchcock, 2022: 65).

In the study by Weller and Monroe (2014:473), respondents who included historians indicated that the most common methods of digital file storage were harddrives, CD roms or cloud-based commercial storage or printing physical versions of text files.

Stressing the importance of research data management for history, Baker (2014) observed that

> *research data generated by an individual historian is at risk of loss if that historian is not able to generate and preserve it in a form they can understand and find meaningful years or decades after the fact, let alone someone else wading through the idiosyncrasies of their research process.*

2.6 Research Data Management Lifecycle Models

Research data management lifecycle models map out the steps and actions of RDM, thereby providing a checklist for researchers so that all steps associated with their research data are identified and carried out (Harvey 2010:34; Higgins, 2012: 18; Carlson 2014: 69).

Cox and Verbaan (2018:26) observed that, while data is created within a research lifecycle, data itself has its own lifecycle going through "a process of creation, cleaning, combination, storage, analysis, and possibly then preservation, sharing and re-use." These processes echo a definition for data management by O'Reilly, Johnson and Sanborn (2012:2) in which 'data management' refers to all aspects of creating, housing, delivering, maintaining, and retiring data.

Numerous data lifecycle models have been developed by various institutions and organisations. Cox and Tam (2018:142) report on the proliferation of lifecycle models; Ball (2012) identifies nine. Hollander et al (2017) noted that there was not a "one-size-fits all solution" to models

and described a number of models for digital data curation and archiving for six Humanities and Social Science disciplines, including history (Hollander, et al, 2017).

Through the application of a data lifecycle, research data becomes an "actual product of research" and an "integral part of research" (Briney, 2015:11). As Di Cresce and King (2017: 229) state, the data lifecycle approach, originally designed for science data, is applicable to humanities data and can provide guidelines for "structuring a data management plan".

Starr (2012:109) notes that data lifecycles usually include five types of activities: planning, collecting, managing, sharing and publishing. Proposals and project plans fall within the planning phase while the collection phase includes fieldwork and data gathering. During the management activity, the data is prepared for long-term curation and deposit into a repository while could include informally making data publicly available (Briney, 2015:9). Publishing is seen as the more formal means of making data available (Starr, 2012: 110) especially where researchers publish both articles and data (Briney, 2015:9). Researchers make a choice about what they consider to be data, what is saved, what is curated, and what is shared, when and with whom, as well as what, when and for how long data is deposited (Borgman, 2015:38).

2.6.1 Research Data Management Lifecycles of Interest to this Study

A number of research data management lifecycle models are of interest to this study. These include the Digital Curation Centre (DCC) Lifecycle Model and UK Data Archive (UKDA) model. A key lifecycle model which forms the framework for this particular study is one created by the University of Hull History Department and which is based on the Digital Curation Centre (DDC) Lifecycle Model. This History lifecycle was the only lifecycle for history identified by Hollander (2018:204). After the identification of these three lifecycles by this researcher, Ma and Xiao (2020) devised a lifecycle model for Digital History, so this has been included in this study. Ma and Xiao (2020:4) observed that in their review of prior research on data practices in digital history, they found very little published work on data practices in digital history.

This researcher took a decision not to include the Open Archival information System (OAIS) Reference Model - ISO 14721:2003 in this literature review, because its focus is on long term preservation for the building of a digital archive. But it does need to be named as it was used by the now-defunct Arts and Humanities Data Service (AHDS) [which was part of the UK Data Archive] upon which to base its work (Dunning, 2006).

2.6.1.1 Digital Curation Centre Lifecycle Model

The Digital Curation Centre (DCC) is an international centre of expertise in digital information curation focussing on building capability and skills for research data management (DCC, n.d). The DCC produced a Curation Lifecycle Model in 2008 which had been developed as a generic idealised framework for conceptualising, planning, implementing and sustaining digital material across the digital lifecycle (Higgins, 2008:135). Because the model is a generic one, it can be applied to any data output from any research domain, but as Cox and Tam (2018: 150) note, it illustrates how data can be "curated with an emphasis on reuse" offering a graphical high-level overview. Pinfield, Cox and Smith (2014: 4) describe the DCC model as "complex, focussing on data curation rather than researcher support".

Figure 2 presents the DCC Curation Lifecycle Model. As Higgins (2008) explained, at the centre of the model are the digital data. The model, which uses concentric circles, describes three sets of activities: Full Lifecycle Actions which support curation throughout the data's lifecycle; Sequential Actions which need to carried out in the correct order to ensure curation; and Occasional Actions which need to be done when the circumstances call for that particular activity (Higgins, 2008: 22).

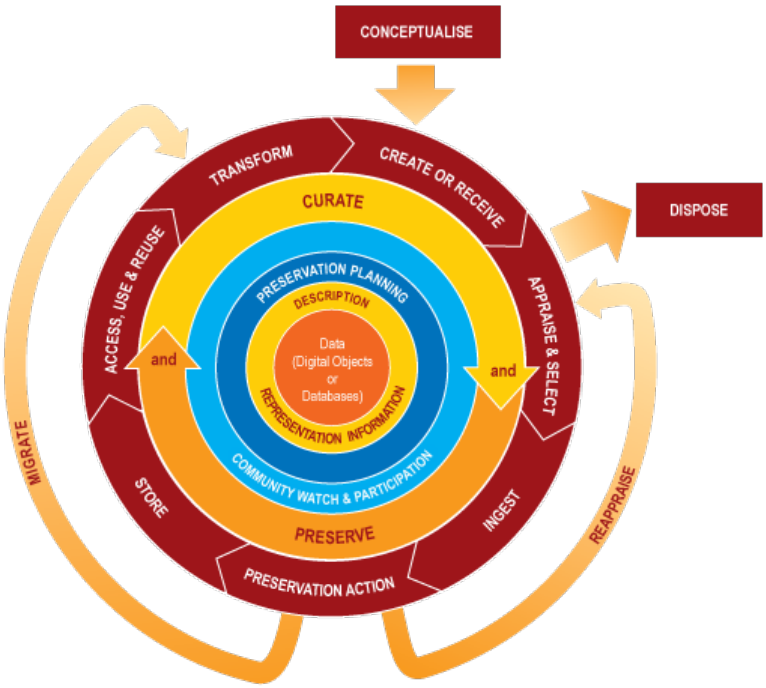

Figure 2: DCC Curation Lifecycle Model (Digital Curation Centre, 2008)

The model is made up of circles each representing a key element or a stage, describing the

various actions of curating and preserving data. The outer circle contains the sequential actions to curate the data, indicating the steps through which the data moves on its journey, that is, from creation, through to transformation and reuse.

The inner circles show the elements that should be considered across the whole lifecycle, while data (in the centre) indicates the object or materials to be curated (Carlson 2014:75). As Johnston (2017: xii) notes, the graphic of the model is quite complex, but the sequential stages and various actions that form the data curation workflow have stood up to the test of time.

The four full lifecycle actions, which support curation through the data's lifecycle, shown as the four inner circles surrounding 'Data' on the model, are 'Description and Representation Information', 'Preservation Planning', 'Community Watch and Participation' and 'Curate and Preserve' (Higgins,2008). 'Description and Representation Information' represents the need for administrative, descriptive, structural and preservation metadata while 'Preservation Planning' not only plans for preservation throughout the lifecycle but includes the administrative and management planning for all the curation lifecycle actions Higgins, 2008). 'Community Watch and Participation' shows the continuous need for the development of and use of appropriate standards and tools, while 'Curate and Preserve' refers to the management and administration planning around curation and preservation throughout the lifecycle (Higgins, 2012).

Sabharwal (2015:95) notes that this lifecycle model does not include "the measuring of the impact of digital curation programmes on the environment", for example, assessing how data collections are used by an institution or community which would be useful for the evaluation of access, use, appraisal and selection activities of the lifecycle.

The relevance of the DCC lifecycle model to this study is that it forms the basis of the History Data Lifecycle model which is described later in this section.

2.6.1.2 UK Data Archive Model (UKDA)

The UK Data Archive Model (UKDA) (Corti et al, 2019) is a data lifecycle model focussing on research data for humanities and social sciences. The UK Data Archive is based at the University of Essex which houses national UK social, economic and population data.

As shown in Figure 3, The model outlines six stages which include creating, processing,

analysing and preserving data, followed by the penultimate stage of giving access to the data followed by using the data.

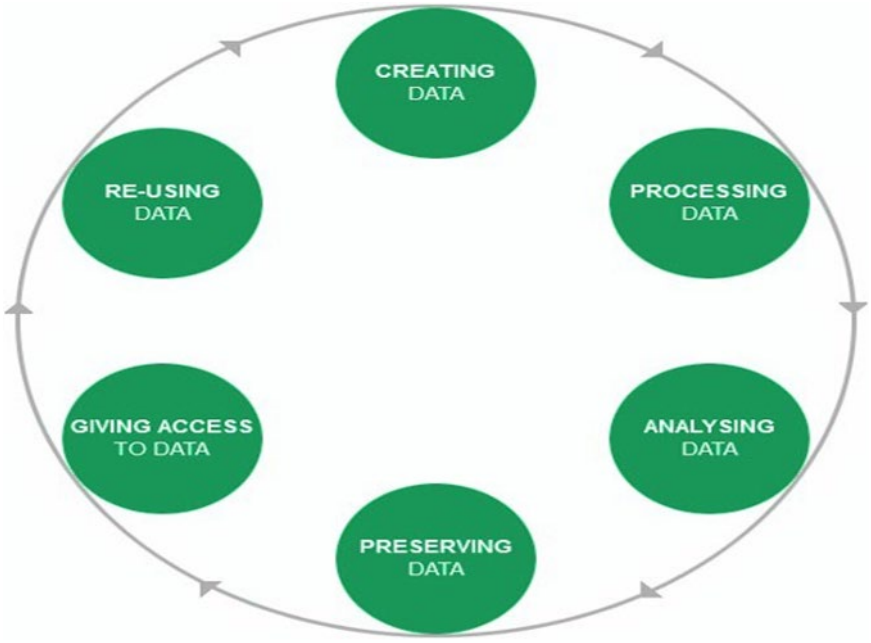

Figure 3: The UK Data Archive research data lifecycle (Woollard and Corti, 2012)

The relevance of this model to this study is that the UKDA hosts the History Data Service which provides access and support for a range of historical databases including promoting and facilitating increased and more effective use of data in research, learning and teaching (History Data Service, 2009). The current collection is listed in the UKDA catalogue. The History Data Service succeeds the Arts and Humanities History Data Service (1996 - 2008) which was part of the UKDA when funding was withdrawn.

2.6.1.3 History Data Management Lifecycle (HDML)

The History Data Management Lifecycle was originally developed by the University of Hull as part of the History DMP Project funded by JISC (Green & Nichols, 2011). It was then further developed by the Institute of Historical Research (SAS), University of Hull and the Humanities Research Institute at the University of Sheffield, funded by the Arts & Humanities Research Council, becoming part of the training provided to postgraduate students through Postgraduate Online Research Training from the School of Advanced Study, University of London (University of London, nd:20). The History Data Management Lifecycle is the only Data Management Lifecycle that is discipline-specific that is key to this study and has only been found in these training materials. As part of the History DMP Project at the University of Hull,

Green and Nicholls (2012:6) reported that there was a need for a history-centric data management plan which followed the outline of the Digital Curation Centre checklist but needed to be less complex.

The History Data Management Lifecycle (HDML) combines the history research lifecycle (HRL) model with the Digital Curation Centre (DCC) model presenting a more history-specific approach to data management, linked to specific data stages in the model. As University of London (nd) notes, the HDML links to the "recognisable elements that the historian will encounter during any research project or process."

In figure 4 below, the outer circle shows the stages of the HDML, while the inner circle reflects the stages of the HRL.

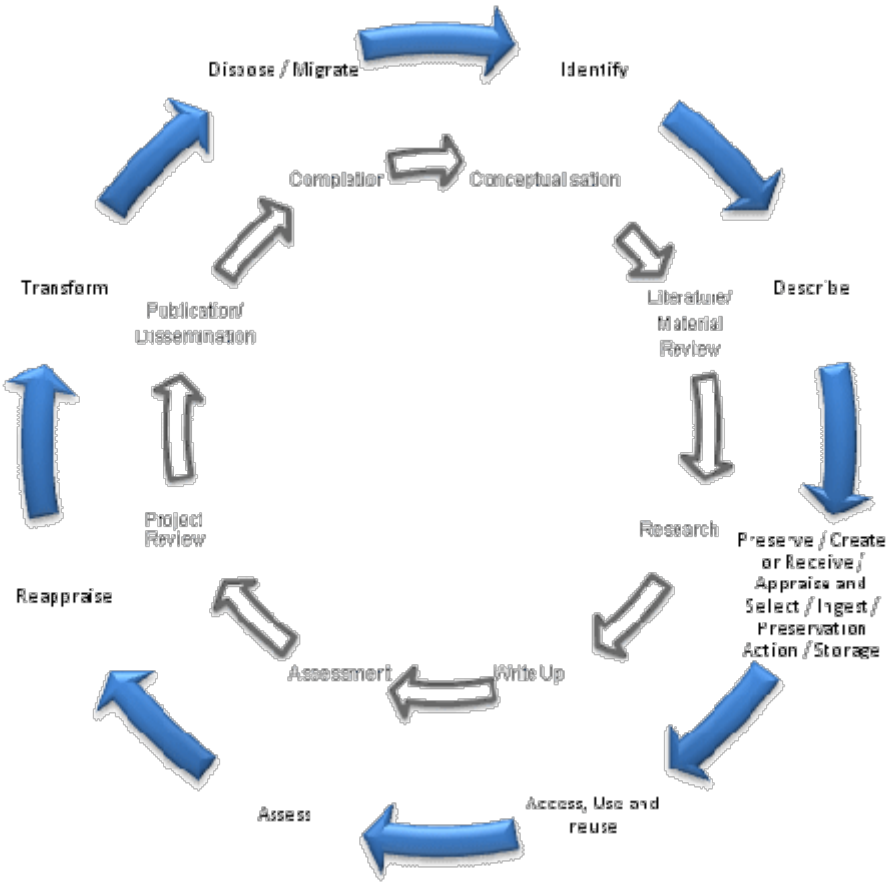

Figure 4: History Data Management Lifecycle mapped to History Research Lifecycle (University of London, n.d)

The "Identify" stage of the HDML is mapped to the 'Conceptualisation' stage of the History Research Lifecycle where the research project is conceived, and would include how its data would be generated, created or incorporated into the research. This is the stage when what is

being researched is identified. This includes the source of that data and its accessibility and the relevance to the research project. Examples of data could be sets of images saved and referenced in research or a transcription of a manuscript.

The ‘Describe’ stage in the HDML is mapped to the Literature Review within the History Research Lifecycle. It is a similar process to that of identifying relevant literature but focusses on the identification of relevant and accurate data needed for the research project. The data is described in full, so that data which assists in the project can be identified. Each data source is referenced accurately, and each element verified. The description includes formats, and the physical dimensions, resolution and storage capacity. Included in the description are copyright and ethical considerations as well as licensing restrictions on use.

The ‘Preserve/Create or Receive/Appraise or Ingest/Preservation/Action/Storage’ Stage incorporates several aspects of the DCC Lifecycle and is mapped to the Research stage within the History Research Lifecycle. These individual stages on the DCC lifecycle have been coalesced into one stage on the HDML as can be seen in Figure 4 which deals with the management of data.

The ‘Preserve’ stage of the data considers and identifies how data will be maintained throughout the research and at the end of the research project. This stage is essential in terms of the availability and accessibility of the data during and after the project. This would include where the data would be stored during and after the project, for example, where the file would be stored and preserved, e.g. on a drive, or server.

‘Create or Receive' refers to new and innovative data that may be created as part of the research process or where new data has been identified that needs to be included in the project. Activities that need to be taken into account include describing the data, how it would be included into the project and what the long-term preservation needs are. An example of an activity related to this stage could be capturing data generated from a logbook into a database to be saved in a usable format.

‘Appraise and Select’ refers to the evaluation of data to determine its validity for the research, including legal, and ethical requirements. For example, if oral history interviews were held, ethical and legal requirements would need to be complied with.

‘Ingest’ refers to the placing or transfer of the data into an archive, institutional or subject

repository or data centre. This enables long-term preservation. The researcher may be required to comply with strict file name processes, checking and verifying the file structures and integrity of the data. (University of London, nd).] 'Storage' refers to where the data is to be securely stored.

The 'Access, Use and Re-Use" Stage in the History Research Data Management Lifecycle relates to the 'Write up' stage in the History Research Lifecycle, and refers to the use of the data, once the data has been managed and is available for use and reuse in the writing up of the project. At this point, the actual writing up is the "remit of the researcher" (University of London, nd) but the research would have been conducted, and the various data management stages as described above, would have been completed.

The 'Assess' stage covers the assessment of the research project which could include identification of further areas of development. This relates to the Assess, Use and Reuse stage of the DCC model. Although this stage in the HDML does not deal with data per se, it is where feedback is received on the project. This could include feedback on data management, for example, in terms of data security. The 'Reappraise' stage follows from the 'Assess' Stage and relates to reviewing of all processes in the research project which could include issues around how data is described and stored (University of London, nd.)

The 'Transform' stage in the data lifecycle, maps to the 'Publication/Dissemination' stage in the History Research Lifecycle. At this point the results of the research are published or disseminated. The 'Transform' aspect may relate to the researcher making the published data available in a variety of formats, for example, via "a database that could be available on the web" (University of London, n.d.) or be placed in a local institutional repository.

The end stage in the History Data Management Lifecycle is 'Disposal and Migration' which maps to the 'Completion' stage in the History Research Lifecycle. In terms of the History Data Management Lifecycle, this refers to the value of the data and its use after the project is completed, which could include disposing of the data or moving it to another location or reformatting to work with newer technologies or overcome issues of obsolescence, where the data would then be available for continuing or new research.

Once the research project has been completed, the data is still available as a "legacy of the completed project" (University of London, nd) and is available for new or continuing projects which leads to the next stage or sequence of events.

The History Department of the University of Hull developed a DMP template for History using this lifecycle (Nicholls, 2012). The template contains notes and guidance about data for the researcher.

2.6.1.4 Digital History Research Lifecycle

In 2021, this researcher found the Digital History Research Lifecycle, created by Ma and Xiao (2020). The study reported on current data practices in digital history research and how these differ from analog history research. (Ma& Xiao, 2020:2).

The lifecycle is very similar to both the UKDA and the History Data Lifecycle and offers a general workflow of digital history research as outlined in the figure below, based on the discussions and survey results in their study (Ma & Xiao,2020:12).

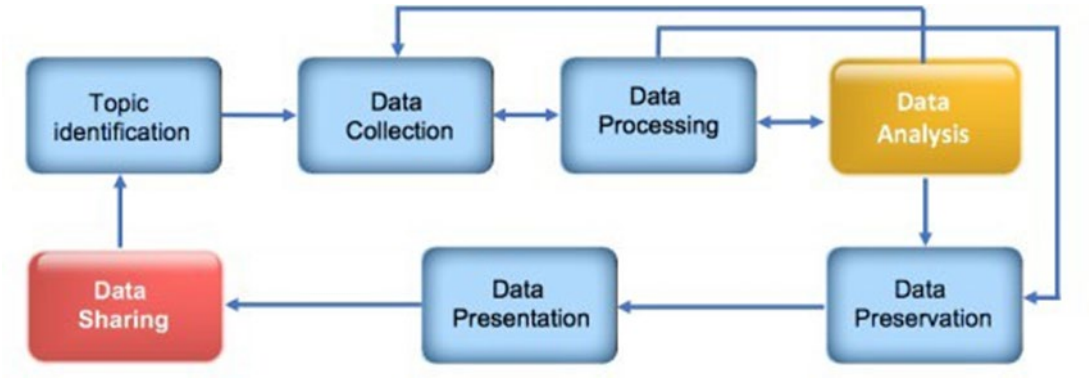

Figure 5: Digital History Research Lifecycle Source: Ma and Xiao (2020)

The lifecycle shows that digital history projects go through stages of data collection, processing, analysis, preservation, presentation and sharing. Ma and Xiao (2020:12) noted that the progression of preservation, presentation and sharing tends to be linear, but the stages from data collection to preservation are iterative where there could be several rounds of "thinking, refining, and working with data". The authors observe that this model does not cover all issues in the digital research process, but stress that it provides the "necessary procedures for fulfilling a digital history project" as well as provide a guide for digital history data practices (Ma & Xiao, 2020:12).

The step 'Data Collection' describes the data collection methods, tools and techniques used in digital history, including traditional archives research (Ma & Xioa, 2021:8). 'Data Processing' refers to the data cleaning, data transformation and de-identification (Ma & Xiao, 2021:9). During the 'Data Analysis' step, computational methods are applied to data for analysis. These

computational methods could be visualisation, text analysis and spatial analysis (Ma & Xiao, 2021:9). 'Data Preservation' refers to the preservation of the final results data as well as the data generated during the processes (Ma & Xiao, 2021:10). 'Data Presentation' speaks to the publication of data through digital platforms such as websites, digital collections, and conference papers and presentations (Ma & Xiao, 2021:11)

The relevance of this lifecycle model to this study is that it is only the second lifecycle found by this researcher specifically discussing History. For this reason, it has been included in the literature review. Although it focuses on digital history research lifecycle, it is relevant as it contains "the necessary procedures of data practices … and serves as a guiding tool" (Ma & Xiao, 2021:12) in the history discipline.

2.7 Support and Services for Research Data Management Required by History Researchers

This section of the literature review describes support and services that researchers, specifically historians, may require from libraries. Numerous studies (Fear, 2011; Jahnke, Asher & Keralis, 2012; Toups & Hughes, 2013; Jones, 2014; McClure et al, 2014; ; Pinfield, Cox and Smith, 2014; Tenopir et al, 2014; Weller & Monroe-Gulick, 2014; Flores, Brodeur, Daniels, Nicholls & Turnator 2015; Mohr et al, 2015; Schopfel & Prost, 2016; Shen, 2016; Hollander et al, 2017; Yu, Deuble, Morgan, 2017; Perrier & Barnes, 2018; Shelly & Jackson, 2018, Sheffield & Burton, 2022; Moulaison-Sandy & Wenzel, 2023) identified research data support and services needed by researchers in the areas of services in particular data management plans, and training and education that libraries could provide.

2.7.1 Services

A range of services are needed to support the creation, management and sharing of research data within an institution. Jones (2013:89) identified the following as elements in that infrastructure: an overarching governance framework to shape the delivery of services; specific infrastructure and services provided at key points in the data lifecycle; and assistance from support staff to encourage the uptake and use of services (Jones, 2013:89). She pointed out that these services are needed at key points in the data lifecycle and include guidance, data management plan templates, tools and consultancy services as well as services during the active phases of the research which could include storage of that data (Jones, 2013:90).

In response to demands from the university or funders, assessment of needs of researchers and staying relevant in the "changing, digital research world", many academic libraries include research data management in the research services they provide (Flores et al, 2015: 82).

A study by Yu, Deuble and Morgan (2017) gave an overview of the University of Queensland Library's Research Data Management Services where the services and support offered were based on the research lifecycle framework. They found that services and support often include Data Management Plan (DMP) consultations as a starting point (Yu, Deuble & Morgan 2017: 289). More advanced or "technical services" may include data curation, preparing data for deposit, creating metadata, or carrying out preservation, and navigating data repositories choices or providing an institutional repository as found by Sheffield and Burton (2022:7).

An earlier study by Weller and Monroe-Gulick (2014:480) reported researchers, overall, identified needing assistance with data analysis, data storage and data dissemination as well as assistance with writing data management plans. However, Shen (2016) found help with short-term storage, long-term storage, and metadata were what faculty wanted more than Data Management Plans consults.

The study by Moulaison-Sandy and Wenzel (2023:81), identified the need for personal digital archiving of data created during research, where researchers were attempting to manage, and preserve a wide range of data. They noted that both the individual researcher and the institution have a responsibility to manage the data created by the researcher (Moulaison-Sandy & Wenzel, 2023: 81).

The Horizon 2020 Parthenos D3.1 Report (Hollander et al, 2017) examined common policies, and attitudes and approaches to data in various disciplines including history, noting that historians seemed to have the most difficulties or "conflictual relationships with recognised and formal policies, mainly as there was not a "concrete need" to develop a data policy" (Hollander, 2017:48) because historians did not create or organise data according to a shared data standard or formal policy. The report noted there was a difference in data creation and organising, with historians making sense of already existing data (Hollander et al, 2017:49). Where historians archived their data depended on the individual researcher or on tacit policies shared by a research group.

2.7.2 Data Management Plans (DMP)

A data management plan (DMP) is a formal document outlining how data will be collected, managed, stored, preserved, shared and disposed of during and after the research cycle. As part of support for RDM, assistance with writing data management plans is often provided by libraries (Yu, Deuble & Morgan, 2017:289).

The study by Weller and Monroe-Gulick (2014:477) reported that researchers whom they surveyed needed assistance with writing data management plans but reported that historians anticipated needing less assistance than other researchers. The study by Berman (2017b:17) reported on perceptions of data management plans noting that, for some researchers, the submission of a data management plan was seen as bureaucratic, "merely as an item to check off a list" with Yu et al. (2017:289) also reporting that it was seen as an administrative burden.

Green and Nicholls (2011:6) observed that a data management plan for historical studies would need to cover the management of data collected via paper sources as well as born-digital materials. Furthermore, it would need to cover data at different stages in its own data life cycle as well as stages in the research lifecycle.

In a study by Berman (2017a:16), researchers reported interest in support for data management plans, data storage and preservation and there was a suggestion that "an informational website with best practices and campus resources" would be useful. Berman (2017a:19) noted researchers could consult this website which would provide "indirect support" for research data management and the need for submissions of data management plans.

2.7.3 Training and Education

For research data management practices to be enabled, and embedded as part of research practices, researchers need to acquire skills to facilitate good practices in managing data, (Corti, 2018; 26). Both Flores (2015:82) and Yu et al (2017:287) reported that the support provided should include training programmes for basic and ongoing awareness and education, through the provision of workshops to suit disciplines or immediate needs.

Schopfel and Prost's (2016) study showed that researchers, including students, needed more information about conditions and opportunities for data management, data deposit and data sharing, including information about policies (2016; 109).

Studies by Fear (2011), Pierun et al (2012), Adamick et al (2013), Johnston & Jeffryes (2014), Carlson & Bracke (2015), Kim and Wong (2015), Wiley & Mischo (2016), Willard (2017), Willaert & Cotton (2019), Ma and Xiao (2020), Smits & Trenkeler (2020), Wiley & Kirby (2020), Sheffield and Burton (2022) explored data management practices amongst graduate students in particular, where the need for RDM support was identified. These studies have generally been broad multidisciplinary studies, rather than focused on Humanities in particular, but all indicate similar challenges where students had "a rather fragmented knowledge of data management" and muddled data management issues with other topics such as methodology and study design (Smits & Treperek, 2020:8). It was suggested that, by focusing on the early career researchers and postgraduate students, the groundwork for good data practices would be laid (Wolski & Richardson, 2015:6).

The graduate students in a study by Wiley and Kerby (2018) focused on RDM in terms of the software they were using, in describing their primary ways to manage the data indicating that their focus was task specific". However, the Rutner and Schonfeld study included graduate students in Historical Studies where students indicated the need to have a "grounding in the methods and practices of the field" (2012: 38) with formal training on research methodologies. Ma and Xiao (2020:14) also observed that the training of historians needed to be updated to include teaching of new knowledge and skill sets including data literacy. This observation is supported by a study of digital humanities training in history undergraduate courses where researchers noted there was a lack of sufficient data literacy and management training (Kim & Wong, 2015:208). These studies point to the need for more resources for the data management education of students by academic institutions, and suggested, regardless of the discipline, there was a need for a comprehensive overview of research data management so students, some of whom are early-career researchers, are trained in good data practices.

Dressel (2017:5) described a workshop presented to new history postgraduate students which focussed on reasons for managing research data, data management plans, file management, documentation, storage, back-up, and long-term preservation. The workshop enabled these students to "contextualise application of the research data management concepts" at the start of their research.

In the study by Ma and Xiao (2020:14), the researchers reported that despite several online programmes such as Programming Historian (which provides tutorials on a range of digital tools), the necessary skills set for research data management have not been integrated into

formal training for historians. They recommend that tailored data practice guidelines for digital historical research are developed which would promote good data practices.

In summary, the support and services required by researchers include a range of services and assistance with data management plans and templates, data analysis, data curation, data storage and preservation and metadata. There was also a need for training programmes for awareness and education.

2.8 Summary

This chapter presented how research data management fits into current research lifecycle practices within historical scholarship, exploring research data lifecycles which could be used by historians. The chapter began by explaining research data and practices in humanities and historical studies. Four data lifecycles which could be appropriate for historical studies were identified. The final section of the literature review explored the support and services that historians could require from the library. The next chapter looks at the research design and methods employed by this study.

Chapter 3: Research Methodology

3.1 Introduction

Research can be defined as a "type of systematic investigation that is empirical in nature and is designed to contribute to public knowledge" (Trochim, Donnelly and Arora, 2016:5). Creswell (2014:5) describes a framework for research in which the three components involved are the philosophical assumptions the researcher brings to the study, the research design or approach related to this philosophy or worldview and the methods or procedures that bring this approach into practice. This chapter describes the study's research approach, design and methods for exploring the research data management practices of researchers in the Department of Historical Studies at the University of Cape Town.

3.2 Worldview and Research Approaches

The broad research approach is the plan to conduct research, and includes the philosophical worldviews or paradigms brought to the study (Creswell, 2014:5). The term 'paradigm' or worldview refers to "a model or framework through which to observe and understand" (Babbie, 2013:57) which shapes the approach to research (Creswell 2014:6).

The paradigm or philosophical worldview best suited for this study is a constructivist paradigm. As Creswell (2014:8) notes, the goal of the constructivist paradigm is to focus on the participants' views of the matter being studied. It seeks to understand the world in which they live and work. The aim is to "rely as much as possible on the participants' view of the situation being studied" (Creswell, 2014:8). For this reason, this is an appropriate philosophical worldview for the study as the participants' views and experiences of research data management practices within the Historical Studies Department will enable the researcher to understand the context and information gathered during the study.

There are three types of research approaches: quantitative, qualitative and mixed methods. The approach that is chosen depends on the research question, in other words, what exactly is being studied, the role of the researcher, and the method of analysis (Punch, 2005:239; Yin, 2014:4; Trochim, Donnelly & Arora, 2016:60).

3.2.1 Quantitative Approach

A quantitative approach collects, analyses, and displays data in a numerical form rather than

narrative. It relies on numbers and statistics in the analysis and interpretation of the findings that are usually extrapolated from the sample to the general population (Bless, Higson-Smith & Sithole, 2013:16).

3.2.2 Qualitative Approach

The qualitative approach refers to "research that investigates aspects of social life which are not amenable to quantitative measurement" (Sumner, 2006: 49). As Given (2008:522) notes, this approach focuses on meanings as conveyed by the participants, which provides a better understanding of the phenomenon under investigation (Bless, Higson-Smith & Sithole, 2013: 16). The qualitative approach is used when the researcher wants to "develop detailed stories to describe a phenomenon", or "achieve a deeper understanding of issues" (Trochim, Donnelly & Arora, 2016: 57).

Methods used in this approach are those which are "open-ended" and "allow the collection of detailed information in a relatively close setting" (Sumner, 2006: 249). These methods include participant observation, in-depth interviewing, discourse analysis, and conversational analysis (Sumner, 2006:249; Trochim, Donnelly & Arora, 2016:60 – 64;). This approach was chosen for this study as it reflects the constructivist worldview and allows for establishing meaning from the views of participants (Creswell, 2014: 19) and as "determining what respondents think and feel about a particular phenomenon or issue" (Bless, Higson-Smith & Sithole, 2013:16).

3.2.3 Mixed Methods Approach

The mixed methods approach has been defined as the combined use of quantitative and qualitative methodologies within the same study to address a single research question (Hewson, 2006:179) with the assumption that this approach would provide a more complete understanding of the research question than using just the quantitative or qualitative approach alone (Creswell, 2014:4).

3.3 Research Design

Research design is the detailed plan of activities to achieve the research objectives (Rwegoshora, 2014:87). Research design has been described as types or strategies of inquiries within the three approaches detailed in the preceding section (Cresswell, 2014:2).

The research design selected for this study is that of a case study. This design is appropriate as the study focuses on the practices of a particular entity: the Department of Historical Studies.

In this case, it is a single case of the current research data management practices in a particular department. The decision to use this approach is a strategic one in particular, as this is a small-scale study with very clear boundaries, focussing on one case to fully understand that particular case, based on this study's research questions.

Thomas (2016:23) defines a case study as an "analysis of a person, event, decision, period, policy, institution or another system that is studied holistically by one or more methods, in other words, enabling one case to be studied in detail, and in its natural setting, enabling a full understanding of the case as possible". A case study enables a researcher to develop a full understanding of a particular case "getting a rich picture and gaining analytical insights from it" (Thomas, 2016:23).

Different types of case studies have been described: intrinsic, where a researcher wants to a better understanding of a particular case, or instrumental, where insights into issues are sought, or collective or multiple case studies, where several cases are covered (Punch, 2005:144), explanatory, where the researcher would be "drilling down" into the topic (Thomas, 2016:123) and exploratory, where little is known and the intent is to "establish the shape of the problem or issue" (Thomas, 2016:132). This small study is an example of an intrinsic and explanatory case study because this researcher aims to have a better understanding of the issue, and to "unpack" or explain the issue.

A characteristic of a case study is that it is required to be a "self-contained entity" and needs to have clear and obvious boundaries," showing what it contained within the case and what is outside" (Denscombe, 2003:38). This case study's boundaries are clearly identified in that the case is that of the research data management practices in the Department of Historical Studies at the University of Cape Town.

There are limitations or disadvantages to using a case study research design. Yin (2014:19) notes traditional concerns about case study research include an "apparent inability to generalise from case study findings" and a lack of rigour which could impact on reliability and validity. Denscombe (2003:39) notes that case studies are often seen as lacking a firm degree of rigour through "producing 'soft' data". This researcher is aware of this perception, so careful attention to detail and rigour has been applied in the use of the case study approach.

A case study is about the particular rather than the general (Thomas, 2016:X), so researchers cannot generalise from case study findings, although as Gerring (2016:30) observes, it should "be possible to put the study into a larger context". Therefore, it can promote understanding or inform practice for similar situations (Leedy, 2015:271). The researcher needs to include sufficient details to enable a reader to make an informed decision on how relevant the findings are to other situations (Denscombe, 2003:37).

3.4 Research Methods

The research methods section provides a description of how a study was conducted, and the rationale for the processes followed (Salkind, 2012:800). This description includes the study population or participants, sampling of the population, data collection, data analysis and data interpretation (Salkind, 2012:800).

3.4.1 Study Population

The study population is the entire group about whom the study wants to draw conclusions (Babbie & Mouton, 2001:100) but it is also possible to "get an accurate conclusion" by studying only a sample of the entire group (Bless, Higson-Smith & Sithole, 2013:62). For the purposes of this study, the population is all research and teaching academic staff members in the Department of Historical Studies at the University of Cape Town, including postgraduate students registered at Masters and Doctoral level in the Department of Historical Studies. There were 12 research and teaching academic staff at the time of data collection, while there were 37 postgraduate students registered at Masters and Doctoral level.

3.4.2 Sampling

There are two types of sampling methods: probability sampling and non-probability sampling. Probability sampling is a method of selecting "large, representative samples" where a sample is randomly selected from a list containing everyone in the population, using probability theory (Babbie,2013:127; Ruane, 2005:109). Non-probability sampling was the more appropriate method to use for this study. There were several non-probability techniques that could be used: purposive or judgmental sampling where a sample is selected on the basis of knowledge of a population, its elements and the purpose of the study (Babbie, 2013: 28); convenience sampling where a sample is built on the basis of finding convenient or available individuals (Ruane, 2005:117); snowball sampling which is built around referrals, for example where an

interviewee is asked to suggest additional people for interviewing (Ruane, 2005:117; Babbie, 2013:129) and quota sampling which is a form of stratified sampling, where there is a predetermined quota of respondents in different categories, which is fixed by the researcher (Leacock, Warrican & Rose, 2009:93; Rwegoshora, 2014:215).

For the purposes of this study, the researchers and academics in the Department of Historical Studies were purposively sampled to ensure that there was a representation of early career, established and emeritus researchers and teaching faculty members. The plan was to approach 6 to 8 researchers from the 12 faculty members in the department, to participate in the study. These researchers had a range of research projects they were working on that would potentially provide rich data for the study. This sampling method was identified as the most appropriate technique on the basis of the researcher's knowledge of the population and the purpose of the study

For the masters and doctoral students in the Department of Historical Studies, convenience sampling was used. This was identified as the most appropriate technique to sample students in the Department of Historical Studies as a means of finding students willing to participate in the study. The students were invited to join a focus group discussion on the basis of their availability and interest in the study. The aim was to have 10 out of the 37 masters and doctoral students who were registered in the department at the time participating in the discussion.

3.4.3 Data Collection Instruments

There are a variety of data collection instruments in qualitative survey research. Interviews, questionnaires and focus groups are some of the main tools of qualitative data collection (Punch, 2005:168; Bless, Higson-Smith & Sithole 2013:187).

For this study, face-to-face semi-structured interviews and focus group interviews were the data collection methods.

3.4.3.1 Interviews

An interview is a tool for seeking information on "individual and personal experiences" from an interviewee about an issue or topic (Hennink, Hutter & Bailey, 2011:109). Interviews have been described as a "purposeful conversation where the interviewer has a set research agenda" (Ruane, 2005:149), but it is not a "two-way dialogue" as the interviewee tells his story, while the researcher's role is to "elicit" the story (Hennink, Hutter & Bailey, 2013:109).

Where in-depth information is required, interviewing is the appropriate method to collect data (Kumar, 1996:115) enabling "individual perceptions, beliefs, feelings and experiences" to be identified (Hennink, Hutter & Bailey, 2011:53).

There are different types of interviews: unstructured, structured and semi-structured. The type of interview selected should be aligned to the research purposes and questions (Punch, 2005: 70). Although unstructured interviews are "useful for exploring a topic broadly", it is difficult to analyse unstructured interview data as each interview is unique (Trochim, Donnelly & Arora, 2016:63). In a structured interview, the researcher asks a list of predetermined questions "and nothing more" (Leedy, 2015:160), in other words, a question-and-answer session, as if a questionnaire was administered orally (Leacock, Warrican & Rose, 2006:99) rather than a discussion on the research topic.

In a semi-structured interview, the researcher includes in the structure or interview schedule a list of issues to be covered which acts as a reminder of what is to be covered during the session (Thomas, 2016:190). The advantage of a semi-structured interview is that the researcher can follow up on points or probe further through the use of open questions. The semi-structured interviews conducted for the purposes of this study were not intended as a question and answer session but rather a discussion so the interview could elicit rich data when the participant is able "to develop ideas and speak more widely on issues raised by the researcher" (Descombe, 2003:67).

Researchers use either an interview guide or an interview schedule as a research tool or instrument for collecting data (Kumar, 1996:109; Ruane, 2005:49). Guides are "relatively unstructured" tools that list the general areas or topics to be addressed while interview schedules are "more structured" listing the key or exact questions to be asked (Ruane 2005: 149).

The questions in the interview schedule (see Appendix B) were based on the questions guiding the study, literature review and the history research lifecycle theory supporting the study and designed to collect data about the current research data management practices of academic staff in the Department of Historical Studies.

3.4.3.2 Focus Groups

Focus group interviews are guided group discussions on selected topics (Ruane, 2005:157)

which is a form of non-directive interviewing (Hennink, Hutter & Bailey, 2011:158) where the researcher interviews several people simultaneously. A focus group discussion is an interactive discussion between pre-selected participants focussing on specific issues (Hennink, Hutter & Bailey, 2011:136). Focus groups collect information on a range of opinions from individuals as well as collecting information on "community perspectives" (Hennink, Hutter & Bailey, 2011: 111,151). The focus group is conducted in an unstructured or semi-structured way, where the researcher draws up a list of broad questions, topics and themes which are then used to develop a discussion among the participants (Bless, Higson & Sithole, 2013:200).

For this study, one focus group interview was proposed. The aim was to have a representation of about 10 masters and doctoral students at different stages of their research in the department participating in the focus group discussion.

The interview schedule for the focus group was designed to collect data about the students' current research data management practices especially in light of the requirement for a research data management plan as part of the registration process as a masters or doctoral student in the Humanities Faculty. As with the semi-structured interviews, the questions in the focus group interview schedule (see Appendix C) were based on the questions guiding the study, literature review and the history research lifecycle theory supporting the study and designed to collect data about the current research data management practices of postgraduate students.

One of the advantages of using a focus group is that the participants can discuss the issues in question with each other, leading to a much deeper insight for the researcher than would have been gained from conducting separate interviews. Although Marshall and Rossman (2016:153) comment that those involved in a focus group are unfamiliar with each other and have been selected because they share certain characteristics relevant to the study question, in this study, because the students are drawn from one department, they were familiar with one another. Hennink, Hutter and Bailey (2011:151) observe that one should be aware of the effect that the levels of familiarity could have on the group discussion.

In addition, this researcher was also the subject librarian for the Department of Historical Studies and was familiar to these participants, so this may have influenced participants' responses. Hennink, Hutter and Bailey (2011:151) suggest that the moderator be a stranger to the participants, echoing the differences between the role of the researcher in a focus group interview to that of moderator or facilitator (Punch, 2005:171). For this particular study, to take

into account these challenges and to mitigate potential observer bias, an experienced academic/researcher who had conducted several focus groups in her own research, facilitated both focus group discussions. The sessions were recorded, and notes taken by this researcher observing the group but sitting "outside the discussion circle" (Hennink, Hutter & Bailey, 2001:155).

3.4.4 Pre-test

Pre-testing is the process used to test the "reliability and validity of new (original survey items)" (Abbott & McKinney, 2013:406) using an "independent small sample of the target population" (Abbott & McKinney, 2013:213). For this small study, the interview schedules for the semi- structured interviews and discussion guide for the focus group, were reviewed by a librarian at UCT Libraries who is familiar with the research area and is also an experienced researcher as well as by a researcher in the Department of Historical Studies who was not included in the sample.

3.4.5 Data Collection

In March 2019, through the Department of Historical Studies, invitations to participate in the research were posted to the department's masters and doctoral students. After a low number of responses, a further two invitations were extended to the students, both by this researcher and the Head of the department, as well using the snowball approach through approaching students, academics and researchers for referrals. Those who responded were provided further information about the study and agreed or disagreed to participate. Originally one focus group was planned for students, but because of logistics and availability on their part, a second focus group was held. While splitting up the students made for focus groups with very small numbers, two focus groups allowed all participants to be interviewed and still allowed participants to take part in a discussion rather than a one-on-one interview. There were three students, one at masters and two at doctoral level in the one focus group, and one from each level in the other focus group. Both sessions, held in May 2019, were facilitated by the same facilitator, using the same interview schedule. Both sessions were recorded and transcribed. Although the number of focus group participants was small, the students were at different stages in their research and year of study, and had widely different research topics, so there was a diversity of perspectives present in both focus groups.

Data was collected by conducting semi-structured interviews with purposely selected

researchers and academics from the Department of Historical Studies. The aim was to have a representation of researchers and academics at different stages in their careers. The invitation to participate in the study was emailed to individual researchers and academics in June 2019. Five recipients responded positively, and interviews were arranged and conducted in July 2019. These interviews took place in the researchers' offices. These interviews were recorded and transcribed.

There were 10 participants in this study as a whole.

3.4.6 Data Analysis

The qualitative data in this study consisted of recordings of, and notes taken during the semi-structured interviews and focus group sessions and transcriptions of these recordings. It is necessary to organise and "classify the data into groups, clusters or classes with similar characteristics" to identify common issues and themes in order to do data analysis (Leacock, Warrican & Rose, 2009:208; Rwegoshora, 2014: 321;).

Preparation of the data involves producing a verbatim transcript of the interview or group discussion and removing identifiers from the data to preserve participant anonymity (Hennink, Hutter & Bailey, 2011:211), providing the "rich detail" needed for qualitative research (Hennink, Hutter & Bailey, 2011:230). This data needs to be "sorted" into topics, ideas and concepts identified in the interview guide and the literature (Bless, Higson-Smith & Sithole, 2013:389; Hennink, Hutter & Bailey, 2011:230).

Transcription of recordings by the researcher is a valuable part of the research because the process brings the researcher closer to the data (Denscombe, 2003:183), but it is time-consuming. Further limitations are that the recorded talk is not always easy to hear, and that people do not always speak in finite sentences, and neither is intonation, emphasis and accents are easy to display (Denscombe, 2003:184). However, these would be supplemented by drawing on the notes taken during the interviews and focus group discussion. The recordings were transcribed by a freelance editor as each interview or focus group session was completed. Before embarking on data analysis, the researcher checked that the transcripts were an accurate reflection of the interviews.

Data were classified into themes which were based on the questions guiding the study, the literature review, the research data management lifecycle, and the history research lifecycle.

These were used by the researcher for analysis of data collected from participants. The researcher coded "by hand" as the relatively small amount of data did not warrant processing by computer software application. Once the data was coded into themes, the researcher used her own analytical ability, using Microsoft Excel for capturing and descriptive analysis purposes, to analyse the content. It was possible to capture the coding and data processing manually as the study was a small one with one case study unit.

3.5 Reliability, Validity and Triangulation

Leacock, Warrican and Rose (2009:104) note that validity describes whether the instrument collecting the data served to answer the research question. They note that if the instrument is invalid, then the data, findings and conclusions would be invalid (Leacock, Warrican & Rose, 2009:104). Bless, Higson-Smith and Sithole (2013:391) refer to external validity as the extent to which research findings can be generalised to a broader population.

Reliability of an instrument refers to that the instrument will consistently measure what it is intended to measure, and that a similar result will be found each time it is administered to a group or similar group (Leacock, Warrican & Rose, 2009:106). However, as Bless, Higson-Smith and Sithole (2013:236) observe, reliability and validity are not always the intention with qualitative research, as the "emphasis is on the understanding of a certain phenomenon".

The four criteria for judging the research quality or trustworthiness of qualitative research are credibility, transferability, dependability, and confirmability (Bless, Higson-Smith & Sithole, 2013:236; Trochim, Donnelly & Arora, 2016:71).

Credibility refers to the credibility of the results of the research showing the findings" make sense" (Bless, Higson-Smith & Sithole, 2013:236). Transferability refers to the degree to which the results can be transferred to other contexts or settings and requires the researcher to give detailed description of the context. Dependability requires the researcher to "thoroughly describe and precisely follow a clear" research strategy which would include, for example, how the data was collected, recorded, coded, and analysed (Bless, Higson-Smith & Sithole, 2013:236). Confirmability requires a critical evaluation of the methodology used and that other researchers should be able to have similar findings by following a similar process (Trochim, Donnelly & Arora,2016:72).

For this study, the data collection instruments were pre-tested, and the interviews and focus

group sessions recorded and transcribed while notes were during the sessions, in order to accurately capture the information. The researcher endeavoured to report accurately to maintain trustworthiness and credibility in their reporting.

Triangulation is the method used to verify and increase the "trustworthiness" of research, where different research methodologies are used to investigate the same phenomenon (Bless, Higson-Smith & Sithole, 2013:238). Thomas (2016:67) describes triangulation with reference to the case study approach as "viewing from several points" rather than viewing from one. This study involves data collection using different methods and different population groups to gather perspectives from several points of view.

Lastly, the data collected from the researchers and academics in the Department of Historical Studies was compared to that collected from the focus group discussion with postgraduate students.

3.6 Ethical Considerations

Research should be carried out in a way that "respects and cares for the participants, maintains integrity in the process and results in studies that are reported honestly" (Trochim, Donnelly & Arora, 2016:34).

Participation in this study was voluntary. All participants were informed of this, and furthermore, were informed that they had the right to refuse participation and could withdraw from the study at any stage.

Anonymity is defined by Babbie (2013:35) when neither the researchers nor the readers of the findings can match a "given response with a given respondent." Although true anonymity is not possible, given the data collection methods of interviews and focus group discussions, participants have not been named in this report.

Confidentiality is when "a research project guarantees confidentiality when a researcher can identify a given person's responses but promises not to do so publicly" (Babbie, 2013:36). In this study, in the interview or focus group discussion, every attempt was made to maintain confidentiality through removing all identifying information and naming participants as Participant A, B, C etc. There was no known harm to participants in this study.

Ethical approval from the Department of Knowledge and Information Stewardship (DKIS) Research Ethics Committee, acting on behalf of the Faculty of Humanities was obtained (Appendix A). University of Cape Town protocols were followed to access the researchers and students in the Department of Historical Studies.

3.7 Summary

This chapter gave a summary of the research approach used by the study. The philosophical worldview was that of a constructivist paradigm, using a qualitative approach. The chapter explained why the case study design was selected for the study. The chapter described the research methods, identifying the study population. The sampling process was also explained, where both purposive and convenience sampling techniques were used. Data collection instruments, which included semi-structured interviews and focus groups, were described. This was followed by discussions of the reliability and validity of the study and the research ethics that needed to be considered for the study to be carried out responsibly.

Chapter 4: Presentation of Findings

4.1 Introduction

This chapter presents the data collected for the study. The objective of the study was to ascertain whether and how the act of research data management is incorporated into the current research practices and historical scholarship of researchers in the Department of Historical Studies at the University of Cape Town, and what support they need from the University of Cape Town Libraries.

Empirical data was collected through semi-structured interviews with researchers and academics in the Department of Historical Studies, and through focus group discussions with Historical Studies masters and doctoral students.

The semi-structured interview schedule for researchers (Appendix B) was divided into three sections: (a) research practices of historians (b) research data management policies and (c) institutional support for RDM. The discussion questions (Appendix C) in the focus group focussed on the research practices of the students using the stages in the research data lifecycle as a guide and ascertaining awareness of the research data management policies at the University of Cape Town.

The findings are presented in the sequence of questions asked.

4.1.1 Participants in this Study

There were 10 participants in this study, five academic staff and five postgraduate students from the Department of Historical Studies. The students who participated in the two focus groups, were either first year or fourth year Masters or Ph.d students each working on their own distinct research project. A description of the focus group participants can be seen in Table 1. They are referred to as Student A, Student B, Student C, Student D and Student E in the report.

Level of Study	Number of Students	Year of Study
Masters Students	2	1 First Year Masters 1 Fourth Year Masters
Doctoral Students	3	1 First Year Ph.d 1 Second Year Ph.d 1 Fourth Year Ph.d

Table 1: Description of Focus Group Participants

Semi-structured interviews were held with the five staff members who were at Senior Lecturer, Lecturer, Associate Professor or Emeritus rank. They are referred to as Interviewee 1, Interviewee 2, Interviewee 3, Interviewee 4 and Interviewee 5 in the report.

Despite the small number of participants in the study, because of their diversity of levels/ranks and research interests, it was felt that sufficient data was collected to be able to address this study's research problem. The study did not seek to generalise its findings.

To further maintain the anonymity of interviewees and focus group participants, the researcher will use the pronoun "they" throughout when reporting on individual participants.

4.2 Presentation of the Findings

The data from interviews with the academics and from focus groups with the students have been merged into the presentation of the findings.

The first part of the schedule for both interviewees and focus groups explored the research projects and the research path followed by the participants in order to determine research methods and how they gathered the resources and information for the research projects.

4.2.1 Current Research Projects

Participants were asked to describe research projects they were currently working on or had recently worked on. The projects that were described are either masters or doctoral dissertations by students or book projects or journal articles by the academics.

The student research areas for their dissertations included the history of the University of the Witwatersrand; the history of the diamond fields in Namaqualand; rural conflict in Lesotho in the late 19th and early 20th century; abortion legislation in South Africa; and an investigation of photographs of a particular event.

A number of researchers from the Department of Historical Studies described book projects in their respective research areas. Topics included trauma theory; a global history of savagery and tribalism in the north eastern frontier of colonial India; the legal government of the British empire in the 19th century; settlers' colonialism and genocide; and a project on race and medical humanitarianism using apartheid as a case study.

From these book projects, journal articles were in progress or had been produced. Researchers have also contributed chapters to edited volumes or have edited volumes themselves.

Researchers have different methods of working towards a completed book, as evidenced by some of their statements:

> *There are a few articles that I am working on with the book project, and around the book project,… which is part of the same research. (Interviewee 4)*

> *I'm an essay writer… So I… mould essays together eventually into a book. (Interviewee 2)*

4.2.2 Research Path

Participants were asked to describe the research path they follow, including providing observations about their sources and methodologies. The participants were asked whether they use primary or secondary sources, whether they sourced them in libraries or in the archives or elsewhere, and whether they were working with digital or print resources.

Research paths vary: some participants work with primary sources such as photographs and newspapers found in archival collections, or with library resources or the collection of oral histories as well as with secondary source material such as journal articles or books. Resources are print or digital resources (which could include oral history interviews or transcripts) or a combination of both.

A number of participants observed that their research question was often sparked by the primary material they found in archives.

> *Often I'll find something in an archive and then I'll get interested in a question, and then I want to answer that question and then I'll go to different archives [to explore that topic further] (Interviewee 1).*

This process was echoed by Student B who noted that one does not know what kind of question to ask until one has looked at the material.

Interviewee 4 observed that their research project was expanding their doctoral research which meant visiting archives and national libraries in various parts of the world.

Different approaches to locating sources, both primary and secondary and both physical and online, were described. All the participants reported a combination of physical visits to archives and libraries, and online access to digital sources.

The students observed that they tended to visit libraries and archives physically and some were planning a sequence of visits to libraries and archives across the country.

As Student C noted,

> *I'm just sort of surveying where I need to go and eventually I will be travelling to some of those places (Student C)*

Interviewee 3 commented that they have to go to multiple archives spread across four continents for their research, although some sources are also digitally available.

Interviewee 3 observed

> *I'm a very archival, archive heavy historian." and the sources were from "different kinds of governmental, inter-governmental, ...standard state archives, but also business companies and business archives (Interviewee 3)*

For Interviewee 4, it is a similar experience with visits to multiple archives and libraries across the world.

On the other hand, Interviewee 5 observed that they mostly used secondary sources:

> *I formulate a project and then go and look for relevant stuff. It's been a long time since I've done the kind of work that I imagine most historians do and that is to find an archive and work through it systematically. .... So I don't spend a lot of time working in the archives because I don't have the time, I draw on the expertise of people like {name and name} who have been sitting in archives here for thirty years. (Interviewee 5).*

As this participant noted

> *I can't go and spend two years in {country} going through documents when somebody else has already done it for me and ... I just need to go and pick out what is needed." (Interviewee 5)*

Interviewee 2 observed that previously, *"... all the articles I wrote were based on some kind of oral history, interviews, a little bit on obviously documentary resources"* but with the current research project, the research path followed *"unusually"* included secondary sources so that the path was *"reading heavy"* (Interviewee 2).

For Student A and Student B, accessing and using published literature, and archival research were the main research path. *"My main tactic for my [research project] has been to look at published literature ... also at newspaper clippings and archival material"* (Student B) Student A indicated that for themself, archival research involving newspapers, newspaper clippings and photographic and television archives was their research path.

4.2.3 Data Collection

This section of the interviews explored data collection where participants were asked what they understood by the term "data" and were asked to briefly describe the data they were collecting, or working with, and what sort of data the research was generating.

All the participants observed that they considered the research materials they were working with to be their research data, and many elaborated further on what they understood their data to be:

> *Data could be anything really. If you're going to use it for the purpose of developing a research project and developing some kind of argument that's grounded in some kind of search gathering, then it's data (Interviewee 2)*

> *When we talk about secondary literature ... you don't tend to think of data, but it is actually data (Interviewee 2).*

Further, Interviewee 2 observed that although "conversations" (one of the methods they use in their research) are not the same as interviews, *"they are a different kind of data."*

> *It's material that can be used to develop analysis, a narrative of what happened*

> *in the past, ... So data would consist of things like archival sources, oral history interviews, transcripts, and newspaper cuttings. Sometimes you get diaries or letters (Interviewee 1)*

The students echoed the definitions given by the interviewees, one of them describing data as the *"raw material that you have to work with on which you base your argument and draw your conclusions"* (Student A). Data is further described as being in different formats *"from a newspaper… to a memo, government publications, ... recorded interviews"* (Student C). Student C commented that even oral interviews which would be transcribed and analysed would be data as well. Another student (Student E) described data as *"the stuff that you've surveyed for the particular project."*

The students elaborated further on what data is, observing, for example, that even the notes that they take in the archive or attach to a photograph or put on to index cards, are data.

> *I would regard data from everything, from a casual remark that I note to a recorded interview or any other primary source. When I look at newspapers, for example, I note all the obvious information such as dates, page numbers, captions, and sizes of photographs. Then there is a section in the notes where I put my own observations, such as the content on the rest of the page. This becomes the data. (Student A)*

Student B agreed with Student A observing that data are *"the things we use to make our arguments"*. One participant commented that secondary literature should be considered data as well, even though one *"did not tend to think of it as data"* (Interviewee 2).

In the follow-up question, which asked about the type of data they were working with, the participants were presented with a list of different types of research data and asked to indicate which traditional and electronic data types they were currently using or had used previously in their research. The data types were identified from a list created by the University of Leicester (2012).

The data types listed were not limited to Historical Studies, instead covered a range of types of data that may be used. These included : Documents (text, Word), spreadsheets; Laboratory notebooks, field notebooks, diaries; Questionnaires, transcripts, codebooks; Audiotapes, videotapes; Photographs, films; Test responses; Slides, artefacts, specimens, samples;

Collection of digital objects acquired and generated during the process of research; Data files; Database contents (video, audio, text, images); Models, algorithms, scripts; Contents of an application (input, output, log files for analysis software, simulation software, schemas); Methodologies and workflows; Standard operating procedures and protocols (University of Leicester, 2012)

Participants in the focus group were asked to mark off the list they were handed, while the interviewees commented on the items on the list. Not all the data types listed above were indicated as used either currently or in the past; in total, 16 formats were selected by at least one person, with 'documents' the most used.

The results are displayed in Figure 6.

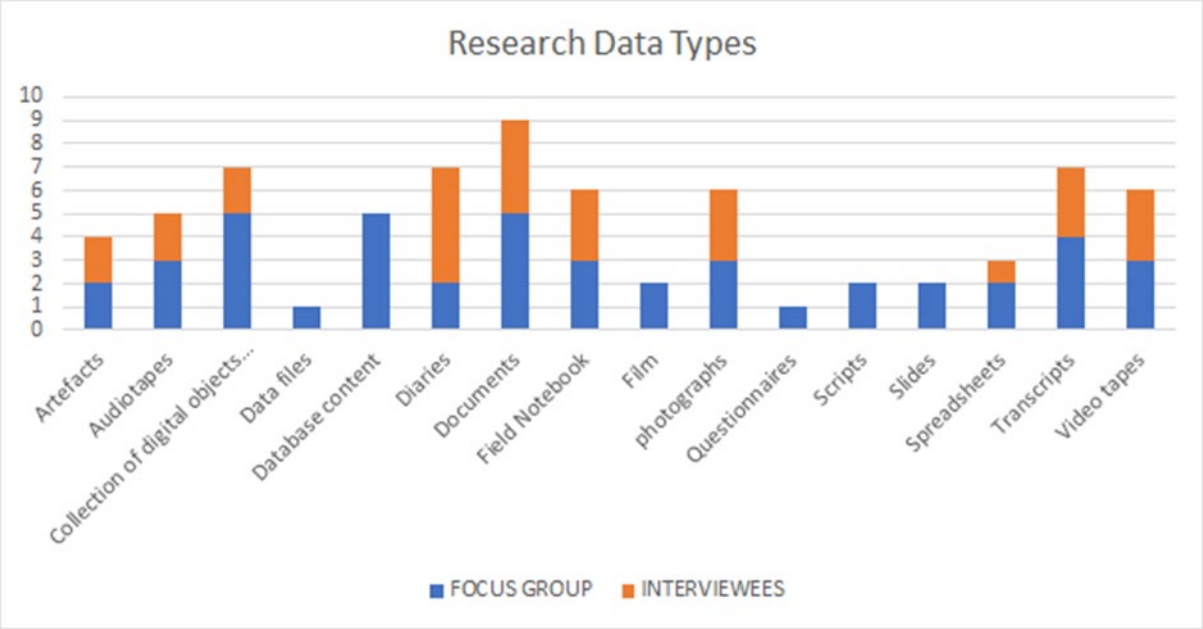

Figure 6: Selection of Research Data Types

4.2.4 Data Curation

Now that the concept of data had been discussed and it was established what participants considered their 'data', the next section of the interview and focus groups asked about primary ways in which the data was managed and organised in terms of capture methods, tools, and documentation related to the data. Participants were asked how they were working with the data during research and keeping track of the research materials and data. A question was also asked about the amount of data collected and whether it would be considered 'big data'.

In general, the participants indicated that data management practices varied depending on the research projects and the data collected. In several cases, the participants did not specifically refer to 'data management' but spoke to managing notes, rather than managing data. Note-taking in the archives is common practice, as is copying and scanning documents, where the archives allow the copying and scanning to happen. "I was trained to take notes in the archives" (Interviewee 3). Interviewee 1 added that they take notes as well, but then work with "an excel spreadsheet just to order my sources … in terms of dates, thematically".

In some cases, the participants were not too concerned about data storage, because much of their data is secondary data:

> *There isn't much need for me to manage or accumulate data and research materials… I don't look to accumulate data. I work with a publication in mind, I conceptualise a project and then I rush through data* [the secondary sources relevant to the project viewed as key resources] *to try to get stuff that is relevant. (Interviewee 5)*

Most of the sources are accessed from the internet and are placed in folders on the computer and are backed up.

> *But I'm not too concerned [about data loss] because I can always go and get it again (Interviewee 5)*

Working with data in the archives was sometimes challenging with limited access to material that could only be used within the archives, and not copied or scanned for analysis on site. Several participants commented on the opening of archives to allow for photography and scanning of material, making these available for analysis outside of the archives, especially where materials had not yet been digitised. This enables this data to be saved and accessed by the researchers so they can work elsewhere with the records.

There are challenges relating to capture of the data. Several participants spoke of the challenges of capturing data from some archives and libraries which relate to the policies and workflows of those institutions. A particular challenge is not being able to take photographs in an archive, so one has to rely on photocopies of the photographs. Another challenge mentioned by Student

A was the length of time it takes to receive the resources requested or to get the print-outs from some archives and libraries. (Student A)

> *"I don't have enough time in the archives … I do also take photos …. I don't have enough time in the archives and … it's very expensive" (Interviewee 1)*

> *I find this* [photographing]*a particularly bad practice because often you do not end up reading all that stuff, they just get stored on your stuff, but if you are writing by hand, you will remember some of it. (Interviewee 3)*

Student B indicated that they use their phone to create digital images, printing these out as hard copies on which they then take notes. In note taking, there is a system which makes sense to them, which may not be obvious to anyone else who reads the notes. They commented that as a history student, there is a lot of scope in developing one's own methodology.

Student A observed that working with images required the best quality copy of the images, so a hard copy was used for quick reference while the digital copy was used for details. Furthermore, Student A also documents all the archival storage devices including the outside of the boxes so there is a record of the archival trace as part of the data collected:

> *I will photograph everything I want to record, ... every article, every magazine cover, every photograph… some of them I print out, some I don't print out, but I sort it all digitally… and code it so that I know which archive it came from and when I took it (Student A)*

> *Because of the delays in receiving the copies, I have to take a lot of notes in advance of getting the copies* [from the archive] *because once they have given you the copy, you don't have access to the newspaper again. (Student A)*

Participants did not specifically comment on metadata and standards although references were made to, for example, coding of photographs or note taking, which implies that there is metadata associated with their data. For example, Student A observed that there was merit in being "very systematic" about coding and recording all information, for example, for newspapers, details like dates, page numbers, captions, sizes of photographs.

None of the participants regarded the data they were collecting as "big data". As Interviewee 3 commented, no one within the Department of Historical Studies Department at the University

of Cape Town currently had any kind of project that could be considered “big data”.

4.2.5 Data Storage

This section of the interview and focus group aimed at determining how data was organised, stored and backed up once the research had been completed.

Almost all the participants report that the data or research gathered is managed on various technological devices or equipment, but that notes are often handwritten. The students, especially, use a variety of platforms including cloud-based platforms and devices such as mobile phones upon which to store their data. Student B preferred to have their resources available online and digitally, also because they work on multiple devices. Several of the participants used platforms such as OneDrive, Google Drive or DropBox was used to store all research materials including interview recordings images and these are backed up (Interviewee 1, Student B and Student C and Student E).

Interviewee 3 describes making notes by hand and retaining those notes as part of data storage, but now relies on the laptop for data storage:

> *I have my older research stuff in piles of hundreds of exercise books, but then once I had the laptop, I just take notes. (Interviewee 3)*

Student D shared that the photographs taken on their phone are transferred to their laptop, but as far as notes were concerned, their preferred method is to “use cards to summarise” and then transfer them to the computer for later organisation. The cards are also organised.

Student C described taking notes in a research notebook*, “as well as using a portable hard drive to back up”* (Student C).

Interviewee 1 noted that they work with *“an excel spreadsheet just to order my sources … in terms of dates, thematically”* but then remarked that they are old fashioned when working with data, in that materials are printed out and *“the whole desk will be filled with different things and then I draw up tables or lists or look for keywords”*. They noted that the print outs were put into lever arch files.

Interviewee 2 saved all data “interviews, the audio and some images and lots of books” on a hard drive, with “loads of handwritten notes”

> *I've used Endnote* [Bibliographic Software] *for my references in my writing which has made a big difference in terms of my ability to work with the cloud across different spaces" (Student B)*

They noted that they were not sure whether Endnote would be regarded as a data management tool. Student A also used Endnote for referencing.

Several participants indicated that the notes they took were organised according to topic, primarily, and then chronologically as one tracks whether there is a change in ideas over a period of time. Student A also regarded the experiences of the consultation of the source as important in their research, so the notes included aspects like keeping notes of the gestures people make during an interview, when doing interviews.

Almost all the participants indicated that backing up of the data was imperative with some reporting losing data previously.

> *At one stage of my career, I was doing the right thing in terms of data management"…'interviews would go into the archives and* [they] *would manage the conservation and the access and all that" …" And now I walk around with this library [on my external hard drive] with it all there and a back up copy on my hard drive. (Interviewee 2)*

> *I have six different drives.. Once I lost a substantial amount of data and after that I just became, actually slightly paranoid. (Interviewee 3)*

Interviewee 1 uses a number of devices such as external and thumb drives. The interviewee reports that backing up is essential having had laptops stolen in the past. *"I'm really dependent on my DropBox for international collaboration and for backing things up"* They also noted that some interviews are on tapes which they should ideally digitise although these have been transcribed in full (Interviewee 1)

I haven't backed up for about 18 months but it's not important … I'm focussed on the writing. (Interviewee 5) but the participant observed that it was important to keep the current writing backed up, especially where several months had been invested in those pages. *"If I lose this, then it is a problem for me."* They noted that they could always get the data again.

Student E observed that they are a *"data hoarder because I realised how you often go back to*

old projects and build on them", keeping them on a computer and on a portable hard drive well. Student C noted that they also use a portable hard drive to back everything up.

4.2.6 Data Preservation

The questions posed in the next section of the focus group and interviews asked about the future of the data once the research has been completed, including about the deposit and long-term preservation of the data.

For the most part, for both students and researchers, the long term prospects for the preservation of data mean keeping the data on storage devices for future publications, with some researchers indicating that they intend to donate their data to the library at some point in the future:

> *I very much still believe in those things of wanting other people to have access to it… and giving it to UCT Library, the archives, you know, but yeah, giving people access to that data (Interviewee 2)*

> *I print it all out and put it into lever arch files and then it's backed up* [in the sense of having a paper copy] *like my transcripts. (Interviewee 1)*

Interviewee 4 described creating their own digital archive through saving archival material they have collected as a back-up and transferring physical hard copies into scans into digital copies for private use.

Interviewee 1 noted that they had thought about their own transcripts and were planning to wait ten to fifteen years and then donate them to the UCT Libraries. Interviewee 5 commented *If there is significant paper stuff, then I gather it together and donate it to our library. … otherwise, I'm going to chuck them out. (Interviewee 5)*

According to Student B *"It's [the data] going to stay on my computer and maybe turn into a few more papers in the future"*. This is echoed by Student A who elaborated further to say the data will eventually be put on a hard drive, while the recordings of interviews will be lodged with UCT Libraries.

4.2.7 Data Sharing

This section explored data sharing in Historical Studies. Questions were posed to determine the practice and attitude towards sharing of data collected by the participants or collected by

other researchers.

There were differing views ranging from a willingness to share to not sharing at all.

> *I share a lot with people who work in my area, in fact, we share all the time. And then there is the more formal sharing through … conferences and symposiums. (Interviewee 5)*

> *I share it* [my data] *personally with the young scholars in the field … I get emails from people, saying, if you have, I just send it to them. Sometimes I ask for something they might send. So I think it is a nice culture we have (Interviewee 3)*

Interviewee 3 commented further to say there has been a perceptual change in academic culture in the history discipline, where people are sharing more than they used to even ten to fifteen years ago. They noted that previously there was an "almost nastily proprietorial" attitude, which was also *"built along particular lines of privilege because not everyone from ex-colonial countries could travel"* to get access to the data. As an example, illustrating the value in sharing data, they commented that by sharing a copy of a document with different people, each person would be reading the document in different ways depending on the particular research questions asked by those individuals and in this way, generating different data to answer their particular research question.

Interviewee 2 however pointed out that data would only be shared at a later stage of research: *"Not while you're writing, but when you're finished"* (Interviewee 2).

Interviewee 1 noted that some data is sensitive and cannot be shared, and *"also I don't really feel people gave me permission to do that* [share their data]*."* The interviewee comments further that other researchers need to go to the archives and follow up on documents used by the interviewee with *"fresh eyes"* and reach their own conclusions. But there was an acknowledgement that use had been made of other historians' data sets, for example, where interviews had been made public.

The students also had mixed feelings about sharing. There are potential benefits of sharing, notes Student B, such as working with people who are doing research in the area. But where they have, for example, collected interviews, they would be uncomfortable with other people using those recorded interviews.

The students indicated a willingness to share their data with affected communities about which they were doing research. Student A was working with family photographs, so all the information collected would be shared directly with the family. Student D had a long-term goal to publish a book so that the information could be shared with the community and fed into official government policies around mining and land reform. They also indicated that photographs they had taken would be given back to the community.

4.2.8 Challenges

The final question in this section related to any ethical or legal challenges or concerns about sharing data, or any other challenge that came to mind.

There were very few ethical and legal concerns expressed around data, although the question of treading carefully in relation to ethics and sharing to benefit communities arose in the discussion, as described in section 4.2.7. Concern was expressed by Student C about the barriers to accessing data from other sources, especially around sourcing and navigating access to the data they need, and the costs to access which were especially frustrating where the records *"are our records"*, that is, belonging to Africa (Student C and Student E).

Student E observed *"as much as you want to share, especially with people who are in your field and to expand the field"* there was a concern about how work is credited especially from overseas institutions interested in Africa, mining the data collected by African researchers who have shared it, or had done the legwork in collecting the data. There was the concern that *"there is not a reciprocal relationship"* leading to reluctance to share the data. Student C observed that, where data had been taken back to archives in the North, there are costs to Global South researchers to access these even though *"these are our records"*. Interviewee 4 commented that making Global South data available *"makes things easier for already privileged institutions and researchers from the north"*. They further comment that these are the *"kind of questions that one has to think through while… thinking about sharing"* but in principle, they supported sharing.

It was observed that history can restore dignity to people who have not been heard. It was felt that, rather than viewing data sharing as administrative or bureaucratic - *"ticking a box for some university committee"* - data sharing was of benefit to the communities in terms of human rights and social justice, *"for altering the world"* (Interviewee 1).

How data was obtained is also an ethical concern. An example cited by a student was about data which had been collected a long time ago through research methods which would be considered unethical now. This data cannot be shared, and other researchers are unable to publish anything that challenges the research previously published from the data because it cannot be done without access to the original data.

Another challenge that was expressed was about materials deteriorating in the archives such as copies of faxes becoming unreadable, or where the ink has faded on letters rending them unreadable. There was an urgent need for digitising these (Interviewee 1).

4.3 Research Data Management Policies

This section related to participants' familiarity or awareness of the University of Cape Town's Research Data Management Policy and its components. As the Research Data Management policy was new at the time of data collection, the intention was to understand how familiar the participants were with the policy. The question about the policy was posed to both sets of participants. Following on from this question, the academics were asked whether they had a DMP for their own research. A follow-up question related to what they shared with their students about DMPs.

The awareness of the official University of Cape Town's Research Data Management policy among participants was mixed. When asked, "Are you familiar or aware of the Research Data Management policy and its components?" responses were:

> *Honestly no.* (Interviewee 1)
>
> *I am aware of it in general, not in detail, no.* (Interviewee 2)
>
> *I did look at it at some point, but it seemed kind of bureaucratically written.* (Interviewee 3)
>
> *I've heard of it. To be honest I have no idea what that policy is.* (Interviewee 4)

The participants in the focus group had no knowledge of the Research Data Management policy, so were unaware that there was a requirement to complete a DMP, but they did indicate that they had seen something in the Memorandum of Understanding which they had signed with their supervisors at the start of the academic year. Student B noted that on the

Memorandum of Understanding signed with her supervisor, they had just said that there was not a DMP.

4.3.1 Data Management Plans (DMP)

Following on from this discussion, researchers were asked whether they had DMPs for their own research. A follow-up question related to what was shared with the students about DMPs.

Very few of the participants knew what a Data Management Plan was or had one. Interviewee 3 noted that: *"It will be interesting to know what to do with all this stuff, but I don't know* [about data management planning]*"* (Interviewee 3).

Interviewee 2 commented that the only plan they had in terms of data was for a research assistant to sort the data that is not in the archive and then hand it over to the library.

Interviewee 1 knew that the National Research Foundation (NRF) required a data management plan but expressed strong opinions against this requirement.

> *I had to put one in for my NRF application … but it just felt bureaucratic and it felt like I'm being forced to do something. To be brutally honest, my judgement as a historian is then being questioned around how I deal with my sources …. I thought that in the methods [section in the application] I'd already covered that. Interviewee 1.*

Interviewee 1 went on to say that a DMP should be tailored to the history discipline.

> *It's like Cinderella, you know like, there's a slipper and we're expected to put our foot into that slipper, as opposed to saying, what about we redesign the slipper. What if we think about what footwear we want.*

Interviewee 1 commented further

> *Human stories are messy… they don't fit onto a spreadsheet, they don't go into little boxes and each story is different and unique… so we need to account for that and develop our own systems because there isn't one way of writing history*

Interviewee 5 responded that if a data management plan was sought after, it would be put together and the data placed in the library.

Interviewee 3 noted that they did not need to adhere to requirements because, *"I tend to shy away from funded projects because I'd rather do research in my own time and own money."*

Student E observed that a data management plan was not something they had been taught *"I think it's stuff you have to like figure it out. ... But I think it should be taught"*.

Interviewee 4 remarked (after a brief explanation from the researcher about a data management plan) that *"in many dissertations, researchers talk about the kind of archive and material they are using so in a sense, a data management plan was basic to any historical research"* but it did seem to be *"just more corporation technical speak"* for something they already do. They continued that *"there was always a plan in terms of how you are going to be generating your data… Any research worth its salt is actually thinking about these ethical, political questions deeply… which is wired in some ways into the project itself."*

4.3.2 Information Shared with Students

The final question in this section asked the participants what information about research data management and data management plans was shared with the students. Since 2019, there has been a requirement for a data management plan in the University of Cape Town's Memorandum of Understanding between the Supervisor and the Student.

The interviewees indicated that there were discussions with students about managing their research, but not specifically about the requirements for a DMP.

Interviewee 3 indicated that the requirement for a data management plan is discussed with the students under their supervision whenever a Memorandum of Understanding is signed. So, if, for example, the student wanted a four-year embargo on their data, it would be the student's decision. *"Research students should be in control of their data… I absolutely go by whatever they want to do and it depends from student to student"*

Interviewee 1 commented that the Memorandum of Understanding was a bureaucratic way of talking about data management but observed that discussions with students were around how to handle sources, backing up and respecting confidentiality. *"Students do need to reflect on their actual method of combing through the data."*

The students in the focus group all commented that they had signed Memoranda of Understanding with their supervisors but were not aware that there was supposed to be a data

management plan. *"None of my supervisors know what a data management plan is supposed to be and so, as a student, I've just said there isn't one, and that's what I've signed off..."* Student B.

One of the students observed that they were not aware of the university's research data management policy and the need for data management plans and was not sure where to start but recognised the importance of having a plan. *"I don't know if it's something we're ever taught ... but I think it should be taught because it's very valuable to think about because the next thing you know you've got 50 electronic books on your computer"*. Student C while Student E commented "*I think it's stuff you have to like figure it out. ... But I think it should be taught".*

4.4 Support from the Libraries

The interviewees were asked about their awareness of the current services and infrastructure provided by the University of Cape Town and the University of Cape Town Libraries to support Research Data Management.

There was very little knowledge about the Research Data Management services and infrastructure that the institution and library offer.

Interviewee 1 indicated *"not really"* to the question about their awareness of the services. Other interviewees expressed similar comments: Interviewee 2 stated, "*Not enough, you know. I'm a bit out of touch."* while Interviewee 5 said *"I look for those things when I need them ... when I run into a brick wall, I might approach someone"* (Interviewee 5).

Interviewee 3 was puzzled by services provided to support RDM, asking *"What are the concrete services?"* They referred to a Research Data Management presentation by the library where the presenter spoke about *"the architecture of data, which was very interesting"* but the presentation did not make clear the connection between history data and library "*... I never really understood the connection between our everyday life and the practical services that the library can provide."* (Interviewee 3).

The final question posed related to how the University and University of Cape Town Libraries could best support the researchers with regards to management of data in an ideal world.

As part of the support for researchers, the need for digital skills training for data management

was expressed by Students A and B. Interviewee 1 said that training should include, for example, information about the range of storage options. The students expressed the need to learn different techniques to save data so that the data can be analysed, for example, through discourse or narrative analysis. For example, Students A and B mentioned that a photograph of a document which is saved as an image file cannot be searched. Student B commented that this training in digital skills is not widely discussed or engaged with in the classes they had attended.

4.5 Summary

This chapter presented the findings of the study in which the research data management practices in the Department of Historical Studies at the University of Cape Town were explored. Data was collected from focus groups and interviews, All participants recognized that the research materials they were collecting were research data. The methods by which this data was captured, managed, stored and preserved varied. Methods for long-term preservation of the data included donating data to the library. There was largely a willingness to share data. There was a mixed awareness of UCT's RDM policies and knowledge of support available from UCT and the library. The next chapter will discuss the findings in terms of the objectives of the study and critical questions, and within the context of the literature review. From this, conclusions will be drawn, and recommendations made.

Chapter 5: Discussion of Main Findings, Conclusions and Recommendations

5.1 Introduction

The previous chapter focussed on the analysis of the data collected through the semi- structured interviews and focus group interviews. This chapter will review the purpose of the study and discuss the findings of the study in an attempt to answer the research questions and achieve the research objectives. Conclusions are drawn and recommendations for future practice and research made.

5.2 Purpose of the Study and Research Questions.

The purpose of this study was an exploration of the research data management practices in the Department of Historical Studies at the University of Cape Town, and of their associated research needs. The research questions asked are as follows: 1. What are the research practices of the researchers in the Department of Historical Studies at the University of Cape Town? 2. To what extent do research data management practices form part of research practices in the Department of Historical Studies at the University of Cape Town? 3. What support for research data management is needed by the researchers in the Department of Historical Studies from the University of Cape Town Libraries?

5.3 Discussion of Findings

Findings are discussed according to the research questions.

5.3.1 Research Practices of the Researchers in the Department of Historical Studies

To understand the research lifecycle of the individual researchers, their research projects and research paths were explored. Several researchers indicated that research questions were often sparked by material found in archives and libraries, instead of research questions coming first, followed by a search for sources. All the participants selected sources (primary or secondary) and explored those sources to identify those which provide relevant evidence.

All participants physically visited libraries and archives across the country or across continents or were planning a series of visits to libraries and archives across the country, but one participant reported that they did not spend much time in archives, but rather drew on the work of people who had been working in archives.It was apparent from the findings that archival work was the principal source for primary sources, echoing Rutner & Schonfeld (2012) who

noted the use of primary sources as key for the historical research method.

The descriptions of the research projects and research paths followed typically matched the elements in the HRL. The HRL shows the stages of research in history, from the identification of a research question to the review of the literature or evidence, to the research phase, followed by the writing up, assessment, review and publication of the project, which is when the project is complete. This is the path generally followed by this study's participants.

5.3.2 Research Data Management Practices as part of Research Practices in the Department of Historical Studies

Participants were asked what they understood by the term 'data' in their research. Despite the literature (RIN, 2011; Schoch, 2013; Borgman, 2015; Posner, 2015; Salo, 2020) referring to challenges in Humanities about what is understood to be data, all the participants observed that the research materials (including secondary sources) with which they were working were data. Participants were able to expand on definitions of data, describing data as "raw material" upon which arguments and conclusions are based.

Clearly the participants in this study understood 'data' in terms of their research and were able to expand and explain what that data was, even describing the notes that they take in archives as data. Like the researchers in Hollander (2017), participants in this study collect, organise and make sense of data that already exists in a library or archive. This was the case with most of the historians interviewed; just one collected data from human subjects using oral history interviews.

Individual participants expanded on their data collection process, describing the data that they were collecting, and the data the research was generating. In the process of this expansion, through describing the actual research process, it was clear that there was an understanding of the benefits and importance of research data management, particularly in the overall history research process and in respect of researchers' own personal data management requirements as described by Corfield & Hitchcock (2022) and Moulisan-Sandy & Wentzel (2023). Although the participants did not refer to 'data management', it was apparent that they were practicing data management.

Data curation practices varied depending on the research projects and the data that was collected. The recording or capturing of content of archival materials is key, so note-taking and

photography (where this is permitted) in the archives remains a common practice as a tool for capturing of the data, but there were challenges experiencing in applying the policies and rules of the different archives. The preparation, organisation and management of research notes is a key element in historical research (McDowell, 2002; Rutner & Schonfeld, 2012). Rutner and Schonfeld (2012) use the term 'research notes management' which is a practice confirmed by several participants who reported using a system of notetaking and managing their notes either manually or digitally. Researchers are using available tools and technologies for managing their data (as per Wolski & Richardson, 2015).

It is apparent that participants have their personalised research processes and data management processes. This reflects the descriptions of data management by Rutner & Schonfeld (2012) and Di Crece & King (2017) who describe a similar process in their studies. Research or data collected are managed on a variety of digital platforms, tools and devices, but that notes would often be handwritten and transferred to the computer. There were diverse methods of digital data storage such as, similarly to the finding of Weller & Monroe-Gulick (2014), external harddrives or thumbdrives or cloud-based storage such as Google Drive or OneDrive. Bibliographic software such as Endnote and Refworks were used to store and manage citations by some participants. Participants indicated that backing up of the data was essential but there was also the suggestion that some resources could be found again relatively easily. Processes varied from participant to participant, so practices were highly personalised and idiosyncratic. Hollander (2017:49) observed similarly that how historians archive or preserve their data depends on the individual historian.

In terms of the future of the data, the participants indicated that, once the research had been completed, the data was kept either in paper form or on storage devices for use in future publications. The printing and saving of physical copies of text or keeping copies of notes digitally or physically is a common practice of preserving data, but methods were individual choices. Rutner & Schonfeld (2012), Sinn & Soares (2014) and Corfield & Hitchcock (2022) describe similar actions. Several participants intended to give their collections to UCT Libraries at some point in the future, which is a form of data management planning even in the absence of a formal plan. This step could be considered a way of making the long-term management of this data a library problem to solve.

Attitudes towards data sharing ranged from a willingness to share to not sharing at all, with some mixed feelings about sharing of data. One interviewee observed that researchers in

history were sharing more than they used to do. There was a willingness on the part of some participants to share their documents where other researchers would be able to extract other data. It was observed how one document could produce different data for different researchers, depending on their research questions. This is echoed by Antonova (2020), although not specifically about sharing documents or evidence, but about recording where data is found so others can locate them.

The student participants raised a number of concerns about sharing data especially when it was shared with researchers from the Global North who may not credit the African researchers or where the data had been taken back to archives in the Global North and not easily accessible. This concern harks back to the report by Akers and Doty (2013) that humanities researchers were reluctant to share data as they were concerned about not being credited for their data. The students expressed readiness to share their data with the communities in which they were carrying out research. This attitude is promising in terms of Research Data Management good practice.

The participants raised few ethical and legal concerns for sharing data, except a stated concern about taking care in handling ethics and sharing. Concern was expressed about how data was available for use now which had been collected through methods which would be considered unethical now. This issue of provenance would mean that the data should not be shared, leaving researchers unable to publish having accessed the original data.

5.3.3 Library Support and Services needed by These Researchers in the Department of Historical Studies in light of the University's Research Data Management Policy

Before determining the library support and services needed for RDM, the awareness of, as well as the attitudes towards the university's research data management policy as well as have to be identified.

5.3.3.1 Research Data Management Policy

The awareness amongst the researchers about the University of Cape Town's Research Management Policy was mixed. Willaert et al (2019:15) observed that the requirements for research data management may be considered an additional administrative "burden" for researchers. Comments from some participants in this study confirmed this, as the policy is seen as a bureaucratic one. It must be noted that there was not much knowledge about the

expectations of the University around RDM and it was not seen as particularly relevant to the participants.

5.3.3.2 Data Management Plans

Very few of the participants knew what a Data Management Plan was or had a Data Management Plan, even though this may be a requirement for the institution, funder, or their discipline. The National Research Foundation (NRF) required a Data Management Plan as part of a funding application which was seen as unnecessary by one of the researchers, as they felt that the plan would be duplicating the information from the research methodology section of the application.

The students were not aware that there was a requirement for a Data Management Plan, but their discussions with their supervisors on managing their research, including handling sources, backing up and respecting confidentiality were certainly about managing research reflecting the practices shared by Williams (2020), Antonova (2020) and Corfield & Hitchcock (2022). As one of the participants commented, in a sense, a data management plan was basic to any historical research.

Participants did seem to have difficulties with the requirement to produce data management plans, reflecting Hollander et al. (2017) who reported that historians seemed to have the most difficulties or "conflictual relationships with recognised and formal policies, mainly as there was not a 'concrete need' to develop a data policy." Some of the participants saw the need for a data management plan as an institutional bureaucratic requirement which could be seen as a box-ticking exercise; one indicated that it was perhaps just easier to comply than not to do so. One of the participants observed that there was a need for a discipline-specific data management plan for historical studies. The History Data Management Lifecycle (see section 2.5.1.3) as developed by Green & Nicholls (2011) could form the basis of a data management plan for historical studies such as that created by Nicolls (2012) for the University of Hull Department of History.

5.3.3.3 Support for RDM needed by Researchers in the Department of Historical Studies from the University of Cape Town Libraries

There was very little knowledge about the institutional support and services from the library. Jahnke, Asher & Keralis (2012) also found in their study that few researchers were aware of

the research data management services a library may be able to provide, though their study was done 10 years ago. There were no recent sources specifying researchers' awareness around RDM services provided by libraries. UCT Digital Library Services offers services which include a data repository and templates for creating Data Management Plans. However, the participants were not fully aware of these services already on offer indicating the need for effective marketing both on the part of the Digital Library Services as well as this researcher as the subject librarian for Historical Studies.

The study by Weller and Monroe-Gulick (2014) noted that historians anticipated needing less help with writing data management plans than other researchers. The study also indicated that digitisation services were in demand by historians (Weller & Monroe-Gulick, 2014) but this was only mentioned by one of the participants in this study, with the others than to acknowledge that materials were available digitally. Hence, the digitisation services of UCT Digital Library Service could be promoted within the Department of Historical Studies.

There are challenges associated with education and training around research data management. A number of student participants expressed a need for training in digital skills that would assist data management, especially in learning various techniques and tools in saving and working with data. Ma and Xioa (2020) listed the lack of technical skills as a barrier to good data management practice, and this seems to be the case in this study. The necessary skills set for research data management have not been integrated into formal training for historians. Ma and Xioa (2020) recommend that tailored data practice guidelines for historical research are developed which would promote good data practices.

5.4 Recommendations for this Study

Having considered the RDM practices of historians and the RDM support they need, the findings indicate that there are a number of strategies and actions that can be considered for implementation in order to improve RDM while still taking into consideration the context of their research practice. There is an opportunity for UCT Libraries and the Department of Historical Studies to cement RDM practices within the Department through these strategies and actions.

There should be a domain-specific strategy towards research data management at UCT as a generic approach does not take the disciplinary differences and culture of history into account, as does the History Data Management Lifecycle (Green & Nichols, 2011), for example. There

should be a recognition that the researchers in the Department of Historical Studies are exercising some research practices in managing their data, using a variety of tools and technologies. There could still be a requirement for a DMP but there should be a data management plan tailored for history available for use, for example, for the Memorandum of Understanding between students and supervisors.

Considering Humphrey (2014) who suggested that the discussion with Humanities (and by extension historians), should be about organising their digital research materials rather than 'managing data', a discipline-appropriate approach would be a better strategy to employ in conversations with the Department of Historical Studies when discussing research data management.

A training strategy focussed on students who are emerging researchers should be devised by the Department of Historical Studies to assist with developing positive attitudes about good data management practices. The recommendation is that data management is incorporated into the history curriculum and the skills integrated into formal training. This formal training could include technical training on digital skills and how to manage digital data. The library could provide support and assist with this training.

Because a number of researchers were planning on donating their research data to the library, the library needs to develop a strategy in preparation of the donations. This strategy could include discussions with the researchers to anticipate the amount and format of the data to be donated and how these would be handled. The donations policy of the library should be updated, if necessary, to take into account any research data donations.

5.5 Study Limitations

A study limitation relates to the sample size in the study. Despite the numerous calls for participants for the focus group, the response was disappointing with only five masters and doctoral students responding. To complicate logistics, they were unable to meet at the same time, hence two separate focus groups were held. Despite this, the discussions in both groups added rich details to the study. The intention was to interview 6 to 8 academics or researchers in the department, to ensure a representation of early career, and established researchers and academics. Only 5 were interviewed due to non-availability because of student protests at the time. Despite this, there are sufficient details in the study to show the relevance of the findings.

5.6 Future Studies

This study focussed on the research data management practices in the Department of Historical Studies at the University of Cape Town. There is very little in the literature related to RDM practices in historical studies, which indicates that there is still much to be explored and added to the body of knowledge. Despite the data collection taking place in 2019, this study is now one of the few studies on the topic.

Recommended future studies should expand the study at the University of Cape Town to include more participants to resolve some of the study limitations. It is also suggested that the study could be expanded to more departments with the Faculty of Humanities for the sake of comparison. Future studies should also explore the research data management practices in other Departments of Historical Studies at other South African universities.

5.7 Conclusion of the Study

This chapter discussed the main findings of the study in relation to the objectives of the study which set out to investigate the RDM practices of the Department of Historical Studies. Three research questions were formulated about the research practices of the researchers in the department. Participants in this study understand that history sources, whether primary or secondary, are research data, but they do not use the term "data" to describe these sources or research materials. The findings show that they are exercising good practice in managing their data as part of the history research lifecycle, although the term "research data management" is not used. Data management practices are part of the historical research lifecycle. The UCT Research Data Management policy is regarded as a bureaucratic requirement, which does not take the disciplinary differences and culture of history into account. There was little knowledge about the services and support available from the University of Cape Town Libraries which is an opportunity for the Department of Historical Studies and the library to work together to cement RDM practices in the discipline at UCT. The study recommends that there should be a domain-specific strategy towards research data management to take the disciplinary differences and culture of history into account, including the tailoring of a data management plan for history at UCT.

References

Abbott, M.L. & McKinney, J. 2013. *Understanding and applying research design*. Hoboken, New Jersey: Wiley.

Adamick J., Reznik-Zellen R.C., Sheridan, M. 2013. Data management training for graduate students at a large research university. *Journal of eScience Librarianship* 2013;1(3): e1022. https://doi.org/10.7191/ jeslib.2012.1022.

Akers, K. & Doty, J. 2013. Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*. 8 (2):5–26. doi:10.2218/ijdc.v8i2.263.

Andrikopoulou,A., Rowley, J. & Walton, G. 2022. Research data management (RDM) and the evolving identity of academic libraries and librarians: a literature review. *New Review of Academic Librarianship.* 28(4):349-365, DOI: 10.1080/13614533.2021.1964549.

Antonova, K, P. 2020. *The essential guide to writing history essays*. New York: Oxford University Press.

Arguing with Digital History Working Group. 2017. *Digital History and Argument white paper.* Fairfax,Virginia: Roy Rosenzweig Center for History and New Media. Available: https://rrchnm.org/argument- white-paper/.

Averkamp, S., Gu, X., & Rogers, B., 2014. *Data management at the University of Iowa: a university libraries report on campus research data needs*. Iowa City,Iowa:University of Iowa. Available: http://ir.uiowa.edu/lib_pubs/ 153/.

Awre, C. 2015. "Research Data Management" as a "wicked problem" *Library Review.* 64 (4/5):356–370. DOI: https://doi.org/10.1108/LR-04-2015-0043.

Ayers, E.L. 1999. *The past and futures of digital history*. Available: http://www.vcdh.virginia.edu/PastsFutures.html [2022, June 6].

Babbie, E. 2013. *The practice of social research*. International 5th ed. Belmont, California: Wadsworth Cengage Learning.

Babbie, E & Mouton, J. 2001. *The practice of social research*. Cape Town: Oxford

University Press.

Baker, J. 2014. *Preserving your research data*. Programming Historian. Available: http://programminghistorian/lessons/preserving-your-research-date.

Ball, A., 2010. *Review of the State of the Art of the Digital Curation of Research Data*. University of Bath. (ERIM Project Document; ERIM Project Document) https://purehost.bath.ac.uk/ws/portalfiles/portal/296172/erim1rep091103ab11.pdf.

Berman, E.A. 2017a. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: findings from the quantitative phase. *Journal of eScience Librarianship.* 6(1):e1098. DOI: 10.7191/jeslib.2017.1098.

Berman, E.A. 2017b. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: findings from the qualitative phase. *Journal of eScience Librarianship*. 6(1):e1097. DOI: 10.7191/jeslib.2017.1097.

Berman E. A. 2017c. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: integrated findings to develop research data services. *Journal of eScience Librarianship*. 6(1):e1104. DOI: 10.7191/jeslib.2017.1104.

Bless, C., Higson-Smith, C. & Sithole, S L. 2013. *Fundamental of social research methods: an African perspective*. 5th edition. Claremont: Juta.

Borgman, C. 2007. *Scholarship in the digital age: information, infrastructure and the Internet*. Cambridge: MIT Press.

Borgman, C. 2012. Conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*. 63(6):1059–1078. DOI: https://doi.org/10.1002/asi22634.

Borgman, C. 2015. *Big data, little Data, no data: scholarship in the networked world*. Cambridge: The MIT Press.

Brennan C. 2018. Digital humanities, digital methods, digital history, and digital outputs: History writing and the digital revolution. *History Compass*. 16:e12492. DOI: 10.1111/hic3.12492.

Briney, K. 2015. *Data management for researchers: organize, maintain and share your data for research success.* Exeter: Pelagic Publishing.

Bulger, M., Meyer, E., Eden, G., Terras, M., Wyatt, S., Jirotka, M., Eccles, K., & Madsen,C. 2011. Reinventing research? Information practices in the humanities: a Research Information Network Report. *SSRN Electronic Journal.* DOI: 10.2139/ssrn.1859267.

Carlson, J. & Bracke, M. 2015. Planting the seeds for data literacy: lessons learned from a student-Centered Education Program. *International Journal of Digital Curation.* 10(1):95-110. DOI: 10.2218/ijdc.v10i1.348.

Carlson, J. 2014. Use of life cycle models in developing and supporting data services. In *Research data management : practical strategies for information professionals*. Joyce M. Ray, Ed. West Lafayette, Indiana: Purdue University Press. 63-86.

Carlson, J.R., 2014. *"How do researchers define their data lifecycle and what can we learn from their definitions?".* Libraries Faculty and Staff Presentations. Paper 46. Available: http://docs.lib.purdue.edu/lib_fspres/46

Chassanoff, A. 2013. Historians and the use of primary source materials in the digital age. *American Archivist*. 76(2):458-480. DOI: https://doi.org/10.17723/aarc.76.2.lh76217m2m376n28.

Cohen, D. J., Frisch, M., Gallagher, P., Mintz, S., Sword, K., Taylor, A. M., Thomas, W. G., & Turkel, W. J. 2008. Interchange: the promise of digital history. *The Journal of American History.* 95(2):452–491. DOI:10.2307/25095630

Consortium of European Social Science Data Archives (CESSDA), 2012. *The Open Archival Information System (OAIS) Reference Model*. Available: https://www.cessda.eu/content/download/496/4465/file/CESSDA%20User%20Guide%20for%20digital%20preservation_2_OAIS.pdf

Corfield, P.J. & Hitchcock, T. 2022. *Becoming a historian: an informal guide*. London: University of London Press, 2022. DOI:10.14296/202205.9781914477164

Corti, L., Van den Eynden, V., Bishop, L. & Woollard, M. 2019. *Managing and sharing research data: a guide to good practice.* 2nd ed. Los Angelos: Sage.

Cox, A & Pinfield, S. 2014. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science.* 46(4):299-316. DOI: https://doi.org/10.1177/0961000613492.

Cox, A, & Tam, W.W.T. 2018. Critical analysis of lifecycle models of the research process and research data. *Aslib Journal of Information Management*. 70(2):142–157. DOI: https://doi.org/10.1108/AJIM-11-2017-0251.

Creswell, J.W. 2014. *Research design: qualitative, quantitative, and mixed methods approaches*. 4th ed. Thousand Oaks: SAGE Publications.

Dalton, M.S & Charnigo, L. 2004. Historians and their information sources. *College & Research Libraries*. 65(5):400–420. DOI: https://doi.org/10.5860/crl.65.5.400.

Denscombe, M. 2003. *Good Research Guide for small-scale social research projects*. 2nd ed. Maidenhead, Berkshire: Open University Press.

Di Cresce, R & King,J. 2017. Developing collaborative best practices for digital humanities data collection: A case study. *College & Undergraduate Libraries*, 24:2-4:226 -237. DOI: https://doi.org/10.1080/10691316.2017.1326330.

Digital Curation Centre. 2008. *Digital Curation Lifecycle Model*. Available: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

Digital Curation Centre. 2010. *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability*. Available: http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf [2023, November 14].

Dreswell, W.F. 2017. Research data management instruction for digital humanities. *Journal of eScience Librarianship.* 6 (2): e-115. DOI: doi:10.7191/jeslib.2017.1115.

Edmond, J. 2016. Will historians ever have big data?. In: Computational History and Data-Driven Humanities. Bozic, B., Mendel-Gleason, G., Debruyne, C., O'Sullivan, D. Eds. CHDDH 2016. *IFIP Advances in Information and Communication Technology*, vol 482. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-46224-0_9.

Fear, K. 2011. “You made it, you take care of it”: data management as personal information

management. *International Journal of Digital Curation.* 2(6):53 -77. DOI: https://doi.org/10.2218/ijdc.v6i2.190.

Flanders, J and Muñoz, T. 2012. *An introduction to humanities data curation. DH curation guide: a community resource guide to data curation in the digital humanities.* Available: https://archive.mith.umd.edu/dhcuration-guide/guide.dhcuration.org/glossary/intro/index.html.

Flores, J.R., Brodeur, J.J., Daniels, M.G., Nicholls, N. & Turnator, E., 2015. Libraries and the research data management landscape. *The process of discovery: The CLIR postdoctoral fellowship program and the future of the academy*. 167:82-102. Available: https://www.clir.org/wp-content/uploads/sites/6/RDM.pdf .

Garwood, J. 2006. Quantitative research. In *SAGE Dictionary of Social Research Methods*, edited by V. Jupp. London: SAGE Publications.

Gerring, J. 2016. *Case study research: principles and practices*. Second edition. Cambridge: Cambridge University Press.

Gibbs, F. & Owens, T. 2012. Hermeneutics of data and historical writing. In *Writing history in the digital age.* K.Nawrotzki & J.Dougherty . Eds. Ann Arbor, Michigan: University of Michigan Press. Available: http://dx.doi.org/10.3998/dh.12230987.0001.001.

Given, L.M. 2008. *The SAGE encyclopedia of qualitative research methods*. Thousand Oaks: SAGE Publications doi: 10.4135/9781412963909.

Gomm, R., Hammersley, M & Foster, P .2009. *Case study method: key issues, key texts*. Thousand Oaks: SAGE Publications.

Graham,S., Milligan, I., & Weingart, S. 2016. *Exploring big historical data: the historian's macroscope*. London: Imperial College Press.

Green, R & Nicholls, J. 2011. *History DMP project: the case for a history data management Plan*. Hull: University of Hull.

Harvey, D. R. 2010. *Digital Curation a how-to-do-it manual*. New York: Neal-Schuman Publishers Inc.

Hennink, M., Hutter, I and Bailey,A. 2011. *Qualitative research methods.* London: SAGE Publications.

Hewson, C. 2006. Mixed Methods Research. In *SAGE Dictionary of Social Research Methods*, V. Jupp, Ed. London: SAGE Publications.

Higgins, S. 2008. The DCC curation lifecycle model. *International Journal of Digital Curation.* 3(1):134-140. DOI: https://doi.org/10.2218/ijdc.v3i1.48

Higgins, S. 2011. Digital Curation: the emergence of a new discipline. *International Journal of Digital Curation.* 6(2):78-88. DOI: https://doi.org/10.2218/ijdc.v6i2.191

Higgins, S. 2012. The lifecycle of data management. In *Managing research data*. G. Pryor, Ed. London: Facet. 17-45.

Hitchcock, T. 2013. Confronting the digital. *Cultural and Social History.* 10(1):9- 23. DOI: https://doi.org/10.2752/147800413X13515292098070

Hollander, H.S., Morselli, F., Admiraal, F., Uiterwaal, F., van Berchum, M., Trippel, T., Offersgaard, L., Di Giorgio, S., et al. 2019. *Parthenos D3.2. Report on guidelines for common policies implementation (Final)*. vol. D3.2. Available: https://pure.knaw.nl/portal/en/publications/parthenos-d32-report-on-guidelines-for-common-policies-implementa [2023, November 13].

Humphrey, C. 2006. *e-Science and the life cycle of research.* Edmonton*:* University of Alberta. Available: http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc.

Jahnke, L.M., Asher, A., & Keralis, S. 2012. *The problem of data: data management and curation practices among university researchers.* Washington, DC: Council on Library and Information Resources. Available: www.clir.org/pubs/reports/pub154/pub154.pdf.

JISC. 2019. *How and why you should manage your research data: a guide for researchers: An introduction to engaging with research data management processes*. Available: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data#.

Johnston, L, et al. 2018. How important are data curation activities to researchers? Gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly*

Communication. 6 (General issue): eP2198. DOI: 10.7710/2162-3309.2198.

Johnston, L. & Jeffryes, J., 2014. Steal this idea: a library instructors' guide to educating students in data management skills. *College & Research Libraries News.* 75(8): 431-434. DOI: https://doi.org/10.5860/crln.75.8.9175.

Johnston, L.R. 2017. *Curating research data volume two: a handbook of current practice*. Chicago, Illinois:Association of College and Research Libraries. Available: http://hdl.handle.net/11299/185335.

Jones, J.A. 1998. Constructing history with computers. In *Writing, Teaching, and Researching History in the Electronic Age: Historians and Computers*. Trinkle, D.A. Ed. Armonk, N.Y.: M.E Sharpe. 83-88.

Jones, S, Pryor, G & Whyte, A. 2013. *How to develop research data management services - a guide for HEIs. DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available: http://www.dcc.ac.uk/resources.how-guide [2015, October 3].

Jones, S. 2013. The range and components of RDM infrastructure and services .In *Delivering research data management services: fundamentals of good practice*. G. Pryor, S. Jones & A. Whyte, Eds. London: Facet Publishing. 89–114. DOI:10.29085/9781783300242.006.

Jordanova, L. J. 2019. *History in practice*. Third edition. London: Bloomsbury Academic.

Key Perspectives. 2010. *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study*. Digital Curation Centre. Available: http://www.dc.ac.za/scarp.

Kim, L and Wong, S. 2015. Discovering the past through data: promoting the design and analysis of original data-sets in history undergraduate courses in Hong Kong. *International Journal of Pedagogies and Learning.* 10(3):208-220. DOI: https://doi.org/10.1080/22040552.2015.1135499.

Kumar,R. 1996. *Research Methodology: a step-by-step guide for beginners*. Thousand Oaks: Sage Publications.

Lavoie, B. 2000. *Meeting the challenges of digital preservation: The OAIS reference model*. Dublin, Ohio:OCLC. Available: https://www.oclc.org/research/publications/2000/lavoie-

oais.html [2023, November 14].

Leacock, C.J., Warrican, S.J., Rose G St. 2009. *Research methods for inexperienced researchers*. Kingston, Jamaica: Ian Randle Publishers.

Leedy, P.D & Ormrod, J.E. 2015. *Practical research: planning and design*. 11th ed. Harlow, Essex: Pearson.

Ma, R., & Xiao, F. 2020. Data practices in Digital History. *International Journal of Digital Curation.* 15(1):1 -22 DOI 10.2218/ijdc.v15i1.597.

Mancilla, H.A., Teperek, M., van Dijck, J., den Heijer, K., Eggermont, R., Plomp, E., der Velden, Y.T. & Kurapati, S. 2019. On a Quest for Cultural Change – Surveying Research Data Management Practices at Delft University of Technology. *LIBER Quarterly.* 29(1):1–27. DOI: 10.18352/lq.10287.

Macdonald, S., Martinez-Uribe, L. 2010. Collaboration to Data Curation: Harnessing Institutional Expertise. *New Review of Academic Librarianship.* 16: 4–16. DOI:10.1080/13614533.2010.505823.

Marshall, C. & Rossman, G.B. 2016. *Designing qualitative research*. 6th ed. Thousand Oaks: SAGE Publications.

McDowell, W.H. 2002. *Historical research: a guide*. London: Longman.

McLure, M., Level, A.V., Cranston, C.L., Oehlerts B. & Culbertson, M. 2014. Data curation: A study of researcher practices and needs. *portal: Libraries & the Academy* 14(2):139-164. DOI: 10.1353/pla.2014.0009.

Milligan, I. 2014. SSHRC's Research Data Archiving Policy and Historians. [Blog, 23 October] Available: http://ianmilligan.ca/2014/10/23/sshrcs-research-data-archiving-policy-and-historians/ [2023, November 13].

Milligan, I. 2019. *History in the age of abundance: How the web is transforming historical research*. Montreal: McGill-Queen's University Press.

Mohr, A.H., Bishoff, J., Bishoff ,C., Braun, S., Storino, C. & Johnston, L.R. 2015. When data is a dirty word: A survey to understand data management needs across diverse research

disciplines. *Bulletin of the Association for Information Science & Technology.* 42(1):51-53. DOI: 10.1002/bul2.2015.1720420114.

Moulaison-Sandy, H., & Wenzel, A.G. 2023. The records data ecosystem in humanities and human sciences scholarship. *portal: Libraries and the Academy* 23(1):67-88. DOI:10.1353/pla.2023.0000.

Munoz, T and Renear, A. 2011. *Issues in humanities data curation*. Center for Informatics Research in Science and Scholarship.Urbana-Champaign: University of Illinois. Available: https://www.ideals.illinois.edu/items/31129 [2023, November 13].

National Endowment Office of Digital Humanities. 2018. *Data Management Plans for NEH Office of Digital Humanities: Proposals and Awards.* Washington DC: Office of Digital Humanities. Available: https://www.neh.gov/sites/default/files/2018-06/data_management_plans_2018.pdf [2023, November 13].

Nicholls, J.H. 2012. *History Department data management plan (DMP): Report*. Hull: University of Hull. Available: https://hydra.hull.ac.uk/resources/hull:5423 [22 June 2022].

O'Reilly, K., Johnson, J., & Sanborn, G. 2012. Improving university research value: a case study. *SAGE Open.* 2(3). DOI: https://doi.org/10.1177/2158244012452576.

Owens, T. 2011. Defining data for humanists: text, artifact, information or evidence? *Journal of Digital Humanities.* 1(1). Available: http://journalofdigitalhumanities.org/1-1/defining-data- for-humanists-by-trevor-owens/ [2021, November 2].

Pennock, M. 2006. Digital Curation and the Management of Digital Library Cultural Heritage Resources. *CILIP.* 25(2):1-6. Available: http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lsl-curation_mep.pdf [12 June 2016].

Pennock M. 2007. *Digital curation: A life-cycle approach to managing and preserving usable digital information.* Available: http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.

Perrier, L., & Barnes, L. 2018. Developing research data management services and support for researchers: a mixed methods study. *Partnership: The Canadian Journal of Library and*

Information Practice and Research, 13(1):1–23. http://dx.doi.org.pallas2.tcl.sc.edu/10.21083/partnership.v12i2.4115.

Pinfield, S., Cox, A M. & Smith, J. 2014. Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE.* 9 (12): e114734. DOI: 10.1371/journal.pone.0114734.

Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P. & LaPelle, N. 2012. Teaching research data management: an undergraduate/graduate curriculum. *JESLIB*, 1(1): 46–50. DOI: https://doi.org/10.7191/jeslib.2012.1003.

Posner,M. 2015. *Humanities data: a necessary contradiction*. [Blog, 25 June] Available: http://miriamposner.com/blog/humanities-data-a-necessary-contradiction/ [2023, November 13].

Pryor, G. 2012. *Managing research data*. London: Facet Publishing.

Punch, K. 2005. *Introduction to social research: quantitative and qualitative approaches*. 2nd edition. London: SAGE Publications.

Ray, J. M. 2014. Introduction to research data management. In *Research data management: practical strategies for information professionals*. J. Ray, Ed. West Lafayette, Indiana: Purdue University Press. 1–21.

Ruane, J.M. 2005. *Essentials of research methods: a guide to social science research*. Oxford: Blackwell Publishing.

Ruediger, D. & MacDougall, R. 2023. *Are the humanities ready for data sharing?* Ithaka S+R. New York: Ithaka S+R. DOI: 10.18665/sr.318526.

Rusbridge, C. & Lavoie, B. 2011. *Draft reference model for economic sustainability of digital curation*. Available: https://unsustainableideas.wordpress.com/2011/09/20/sustainability-strategy-revised/.

Rutner, J. & Schonfeld, R. 2012. *Supporting the changing research practices of historians: final report from IthakaS&R*. Available: http://www.sr.ithaka.org/wp-content/uploads/2015/08/supporting-the-changing-research-practices-of-historians.pdf [2023, November 13].

Rwegoshora, H.M.M. 2014. *A guide to social science research*. Dar Es Salaam: Mkuki na Nyota.

Sabharwal, A. 2015. *Digital curation in the digital humanities : preserving and promoting archival and special collections*. 1st edition. Amsterdam, Netherlands: Chandos Publishing.

Salkind, N. J. 2012. *Exploring research*. 8th ed. Boston: Pearson.

Salo, D. 2020. Is there a text in these data? the digital humanities and preserving the evidence. In *Reassembling scholarly communications: histories, infrastructures, and global politics of open access*. Eve, M., & Gray, J. Eds. Massachusetts, Cambridge: The MIT Press. DOI: https://doi.org/10.7551/mitpress/11885.003.0023.

Scaramozzino, J.M., Ramírez, M.L. & McGaughey, K.J., 2012. A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries.* 73(4):349-365. DOI: https://doi.org/10.5860/crl-255.

Schoch, C. 2013. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*. 2(3):2-13. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy- data-in-the-humanities/.

Schöpfel, J. & Prost H. 2016. Research data management in social sciences and humanities: A survey at the University of Lille (France). *Libreas: Library Ideas.* 29:98-112. Available: http://hal.univ-lille3.fr/hal-01395816/document.

Sheffield, M. & Burton, K.B. 2022. Research Data Management needs assessment of Clemson University. *Journal of Librarianship and Scholarly Communication.* 10(1): eP13970. DOI: https://doi.org/10.31274/ jlsc.13970.

Shelly, M. & Jackson, M. 2018. Research data management compliance: is there a bigger role for university libraries? *Journal of the Australian Library and Information Association.* 67(4):395-410. DOI: https://doi.org/10.1080/24750158.2018.1536690.

Shen, Y. 2016. Strategic planning for a data-driven, shared-access research enterprise: Virginia Tech research data assessment and landscape study. *College and Research Libraries*. 77(4):500-519. DOI: https://doi.org/10 .5860/crl.77.4.500.

Sinn, D. & Soares, N. 2014: Historians' use of digital archival collections: the web, historical

scholarship and archival research. *Journal of the Association for Information Science and Technology.* 65 (9):1794–1809. DOI: http://doi.org/10.1002/asi.2309.1

Smits, D.A.B. & Teperek, M. 2020. Research data management for master's students: from awareness to action. *Data Science Journal.* 19(30):1-11. DOI: http://doi.org/10.5334/dsj-2020-030.

Starr,J., Willett,P., Federer,L., Horning, C., & Bergstrom, M.L.. 2012. A collaborative framework for data management services: the experience of the University of California. *Journal of eScience Librarianship* 1(2): e1014. DOI: http://dx.doi.org/10.7191/jeslib.2012.1014.

Sumner, M. 2006. Qualitative research. In *SAGE Dictionary of Social Research Methods*, edited by V. Jupp. London: SAGE Publications.

Tenopir,C., Sandusky, R., Allard, S., & Birch, B. 2014. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* .36 (2): 84-90. DOI https://doi.org/10.1016/j.lisr.2013.11.003.

Ter Braake, S., Fokkens, A., Ockeloen, N., & van Son, C. 2016. Digital history: towards new methodologies. In *Computational history and data-driven humanities. CHDDH 2016. IFIP Advances in Information and Communication Technology*, vol 482. Bozic, B., Mendel-Gleason, G., Debruyne, C., O'Sullivan, D. Eds. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-46224-0_3.

Thomas, G. 2016. *How to do your case study*. 2nd edition. Thousand Oaks.: SAGE Publications.

Thoegersen, J.L. 2018. "Yeah, I guess that's data": data practices and conceptions among humanities faculty. *portal: Libraries and the Academy*. 18(3):491-504. DOI: https://doi.org/10.1353/pla.2018.0030.

Toups,M. & Hughes, M. 2013. When data curation isn't: a redefinition for liberal arts universities. *Journal of Library Administration.* 53(4):223-233, DOI: 10.1080/01930826.2013.865386

Trinkle, D.A. 1998. *Writing, teaching and researching history in the electronic age:*

Historians and Computers. New York: M.E. Sharpe.

Trochim,W., Donnelly, J P, & Arora, K. 2016. *Research Methods: The essential knowledge base.* 2nd edition. Boston: Cengage Learning.

Uva, P. 1977. *Information-gathering habits of academic historians: report of the pilot study*. Syracuse: State University of New York. Available: https://files.eric.ed.gov/fulltext/ED142483.pdf

University of Cape Town. 2018. *University of Cape Town Research Management Policy.* Available: https://uct.ac.za/sites/default/files/content_migration/uct_ac_za/39/files/TGO_Policy_Research_Data_Management_2018.pdf

University of Leicester. 2012. *Research Data - Definitions*. http://www2.le.ac.uk/services/research-data/documents/UoL_ReserchDataDefinitions_20120904.pdf

University of London. Nd. Postgraduate Online Research Training (PORT). History. London: University of London. Available: https://port.sas.ac.uk/mod/subpage/view.php?id=1837 [2022, June 15].

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. & Horton, L. 2011. *Sharing your data - why and how: managing and sharing data*. Available: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf [2023, November 13].

Weller, T. & Monroe-Gulick, A. 2014. Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*. 32(3):467–482.

Weller, T. & Monroe-Gulick, A. 2015. Differences in the data practices, challenges and future needs of graduate students and faculty members. *Journal of eScience Librarianship*, 4 (1):e1070. DOI: http://dx.doi.org/10.719/jeslib.2015.1070.

Wengraf, T. 2001. *Qualitative Research Interviewing*. Thousand Oaks: SAGE Publications.

Whyte, A. & Tedds, J., 2011. *Making the case for research data management*. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available: https://www.dcc.ac.uk/guidance/briefing-papers/making-case-rdm.

Willaert, T., Cottyn, J., Kenens, U., Vandendriessche, T., Verbeke, D. & Wyns, R. 2019. Research data management and the evolutions of scholarship: policy, infrastructure and data literacy at KU Leuven. *LIBER Quarterly.* 29(1):1–19. DOI: http://doi.org/10.18352/lq.10272.

Williams, P., Stevenson, I., Nicholas, D., Watkinson, A. & Rowlands, I. 2009. The role and future of the monograph in arts and humanities research. *Aslib Proceedings.* 61(1): 67-82. DOI: https://doi.org/10.1108/00012530910932294.

Wiley, C.A. & Kerby, E.E. 2018. Managing Research Data: Graduate Student and Postdoctoral Researcher Perspectives. *Issues in Science and Technology Librarianship*. 89, Spring. DOI:10.5072/FK2KH0P44Z

Williams, R.C. 2020. *The Historian's toolbox: a student's guide to the theory and craft of History*. 4th edition. New York: Routledge.

Wittenberg, J., Sackman, A. & Jaffe, R. 2018. Situating Expertise in Practice: Domain-based Data : Management Training for Liaison Librarians. *Journal of Academic Librarianship*. 44: 323-329. DOI: https://doi.org/10.1016/j.acalib.2018.04.004

Wolski, M. & Richardson,J. 2015. *Improving Data Management Practices of Researchers by using a Behavioural Framework*. THETA: The Higher Education Technology Agenda. Gold Coast: 11 - 13 May.

Woollard, M. & Corti, L. 2013. Case study 4: A national solution – the UK Data Service. In *Delivering Research Data Management Services: Fundamentals of Good Practice*. G. Pryor, S. Jones, & A. Whyte, Eds. London: Facet Publishing. 191 – 204. DOI:10.29085/9781783300242.010.

Yin, R.K. 2014. *Case study research: design and methods*. 5th ed. Thousand Oaks: SAGE Publications.

Yu, F., Deuble, R. & Morgan, H. 2017. Designing research data management services based on the research lifecycle: a consultative leadership approach. *Journal of the Australian Library and Information Association.* 66(3), 287-298. DOI: 10.1080/24750158.2017.1364835.

Appendix A

Library and Information Studies Centre
University of Cape Town
Upper Campus

Private Bag X1, RONDEBOSCH, 7701 South Africa
Level 6 Hlanganani, The Chancellor Oppenheimer Library
Tel: +27 (0) 21 650 4546 Fax: +27 (0) 21 650 2529
E-mail: lisc@uct.ac.za
Internet: www.lib.uct.ac.za/lisc

Ref No.: UCTLIS201810-16 22 October 2018

Ms Ingrid Thomson,
Library and Information Studies Centre,
University of Cape Town.

Ethics Approval for Master's Research

Dear Ms Thomson,

I am pleased to inform you that on behalf of the Faculty of Humanities ethics clearance has been granted by the Ethics Review Committee of the Library and Information Studies Centre for you to proceed with collecting data for your Master's study entitled '*Understanding the Research Data Management Practices of Researchers in the Department of Historical Studies at the University of Cape Town*'.

You are reminded to seek further permissions from the proposed data collection site before commencing with your data collection.

I wish you well with your data collection and the completion of your research.

Yours sincerely,

Mr Patrick Mapulanga
Chair: Department (LISC) Research Ethics Committee

Appendix B

Semi-Structured Interview Schedule for Researchers and Academics in the Department of Historical Studies, University of Cape Town

My name is Ingrid Thomson and I am a master's student in the Library and Information Studies Centre at the University of Cape Town. I am carrying out a research study to Investigating the current research data management practices in the Department of Historical Studies at the University of Cape Town.

The objective of the study is to understand where and how the act of research data management is incorporated into the current research practices and perspectives of historical scholarship within the Department of Historical Studies at the University of Cape Town, particularly in light of the institution's new Research Data Management policy. This will assist in formulating an approach for librarians to providing and enhancing support for research data management within Historical Studies.

My study is being supervised by Michelle Kahn.

You have agreed to participate in this interview by completing the Informed Consent Form and returning it to me. Your responses will be treated with utmost confidentiality. I will ensure anonymity of interview participants by naming participants A, B, C, etc. Thank you for accepting to be part of this research. You are at liberty to withdraw at any point. I would like to confirm your agreement that the interview be recorded.

The interview schedule is in three sections:-

Research practices of Historians

1. Tell me about a research project that you are currently working on or have recently worked on

2. Describe your research path for me. *Aim is to determine research methods and capturing of information e.g. how information is captured in the archives or in the use of digital collections.*

Data Collection (Create, select, appraise):

3. What do you understand by data? *This could be termed "research materials".*

4. Briefly describe the data or materials you collect for your research. What sort of data are you working with? What sort of data is this research generating? *Show list of data types; attempting to determine is understood by data in terms of historical studies. Storage of data or back up of data? Is the research data that you are using primary and secondary sources? Analogue or digital?*

Data Curation (Ingest. Access and Re-use)

5. What are the primary ways that you currently manage the data and research materials you have gathered. *Tools; documentation related to the data; project management principles*;

6. Approximately how much research data are you collecting? Would it be considered "big data"? *(megabyte, gigabyte, terabyte ...)*

Data Storage

7. How are you working with your data/research notes while you are working on your research? How do you organise your data while you are in the active phase of your research and once you have completed the research? *Structuring of data, formats, standards and capture methods*

8. Do you back up your data while collecting and analysing? Is the data secure/safe?

Data Preservation

9. What happens to the data once you have completed the research? What are the long term prospects for the data? *Looking at deposit and long term preservation of data. (leads on to the question of post research repository).*

Data Sharing

10. In historical studies, do researchers share their datasets? Have you used other historians' datasets? *Determining the practice in the discipline about sharing.*

11. Would your data be reusable by others? Would you share data, and with whom?

Challenges

12. Do you foresee any ethical or legal challenges or concerns to sharing data? Does the data you collect or work with have copyright or proprietary? *(Opportunity to identify attitude to data management)*

Research data management policies

- Are you familiar/aware of the Research Data Management policy and its components at UCT
- Does your funder (if there is one) require a data management plan
- Do you have DMP plans for your own research? Does your institution, funder or discipline or research community require you to: plan for the management of the data you will collect/produce?
- Does your institution, funder or discipline or research community require you to share data?
- What do you tell your students?

Institutional including Library support

- Are you aware of current services and infrastructure provided by the University and Library? Have you made use of them?
- In an ideal world, how can the University and Library best support?

Appendix C

Focus Group Discussion Information Sheet and Checklist

My name is Ingrid Thomson and I am a master's student in the Library and Information Studies Centre at the University of Cape Town.

I am currently undertaking research towards my Master of Library and Information Studies at the University of Cape Town in the Library and Information Studies Centre. The topic of my study is "Understanding the Research Data Management Practices of Researchers in the Department of Historical Studies at the University of Cape Town".

The objective is to understand where and how the act of research data management is incorporated into the current research practices and perspectives of historical scholarship within the Department of Historical Studies at the University of Cape Town, particularly in light of the institution's new Research Data Management policy. The study will assist in formulating an approach for librarians to providing and enhancing support for research data management within Historical Studies.

My study is being supervised by Michelle Kahn (michelle.kahn@uct.ac.za).

You have agreed to participate in this focus group discussion by completing the Informed Consent Form and returning it to me. Your comments will be treated with utmost confidentiality. I will ensure anonymity of focus group participants by naming participants A, B, C, etc. Thank you for accepting to be part of this research. You are at liberty to withdraw at any point. I would like to confirm your agreement that the interview be recorded.

I would like to remind you that there is no right or wrong response in this discussion. I am interested in knowing what each one of you thinks about the issues at hand so feel free to be frank in sharing your views regardless of whether it is opposed to the view held by other participants. Your opinions matter and it is important that you state them without fear.

Most likely, you will not like your views expressed to be repeated outside this forum. Kindly treat the views of others the same way and maintain utmost confidentiality of the discussions. My colleague, Susanne Noll, will take on the role of facilitator for this focus group discussion.

Discussion theme 1: Research practice

Discussion points:

1. Discuss their research projects that the students are currently involved in. The students are likely to have their own thesis topic or are working on a departmental project like History Access or working as a research assistant for one of the academics.

2. Ask them to discuss their research paths whether using primary or secondary resources available in the library or archives or elsewhere, and whether these are print or digital, and how this information is captured. Aim is to determine research methods and capturing of data e.g. how data is captured in the archives or in the use of digital collections. At this point, no mention of the word 'data' as that comes into the next question. These two questions are to elicit an understanding of the research path.

Discussion theme 2: Data sources

Discussion points:

3. Ascertain what is understood by "data" or "research data" in terms of historical studies and whether other terminology is used e.g. research notes.

Discussion theme 3: Data types

Discussion points:

4. Using this Information Sheet describing different types of "data", ask them to identify the type of data their research is generating and what sort of "data" they are working with. (Allow time for the students to read through the Information Sheet)

Discussion theme 4: Data storage and preservation

Discussion points:

5. Describe how you are working with your research notes/data while working on your research; how are you keeping track of the research materials and data you have gathered?

6. Identify the storage media and tools you are using Looking at tools they use for

capturing the data, formats, standards, storage of this data. One study refers to this as research notes management.

7. Ask the students what will happen to the data once they have completed the research. What are the long-term prospects for the data? Will they be useful in the long-term? Looking at deposit and long-term preservation of the data.

Discussion theme 5: Data sharing

Discussion points:

8. Would your data be reusable by others? Would you share your data, and with whom? Determining the practice in the discipline about sharing.

Discussion theme 6: Data needs

Discussion points:

9. UCT is now asking for a Research Data Management plan as part of the MOU between the supervisor and student as from 2019. What do they understand by a Research Data Management plan?

10. Are you aware of any services provided by the institution, including the library, to support you with your data management needs? If no response, the facilitator provides examples of some RDM services to see whether some response is forthcoming.

For the facilitator:

- Summarising main points.

- "Is there anything else, pertaining to the subject of this discussion that you would like to highlight, ask or comment on, before we finish the session?"

- Thank participants for their time and contribution.

CHECKLIST OF DISCUSSION THEMES:

Research Practice	
Data Sources	
Data Types	
Data Storage and Preservation	
Data Sharing	
Data Needs	