THE RELATIONSHIP BETWEEN RESEARCH DATA MANAGEMENT AND VIRTUAL RESEARCH ENVIRONMENTS

by
Barend Johannes van Wyk
01235346

Submitted in fulfillment of the requirements for the degree
DPhil Information Science

in the

DEPARTMENT OF INFORMATION SCIENCE,
FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND
INFORMATION TECHNOLOGY

at the

UNIVERSITY OF PRETORIA

Supervisors:
Prof TJD Bothma
Dr MA Holmner

February 2018

Declaration Regarding Plagiarism

I (full names & surname)	Barend Johannes van Wyk
Student number	01235346
Topic:	The relationship between Research Data Management and Virtual Research Environments

Declare the following:

- I understand what plagiarism entails and am aware of the University's policy in this regard.
- I declare that this thesis is my own, original work. Where other people's work was used (whether from a printed source, the Internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
- I have not used work previously produced by another student or any other person to hand in as my own.
- I have not allowed, and will not allow, anyone to copy my work with the intention of presenting it as his/her own work.

25 February 2018

Signature

Date

EDITING DECLARATION

I hereby declare that I have edited the entire language content and technical appearance of this study.

M.S. van Wyk	25 February 2018
Signature	Date

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to:

- Professor Theo Bothma and Dr. Marlene Holmner for their invaluable advice, direction, leadership and input as study leaders;
- Dr. Heila Pienaar for her support and advice during this study;
- My loving wife Anna-Mart van Wyk for her encouragement, patience, support, prayers, and also for her assistance in editing and proofreading the thesis;
- All my friends and family for their support and prayers;
- My heavenly Father for granting me the necessary knowledge, insight, perseverance and strength.

THE RELATIONSHIP BETWEEN RESEARCH DATA MANAGEMENT AND VIRTUAL RESEARCH ENVIRONMENTS

ABSTRACT

The aim of the study was to compile a conceptual model of a Virtual Research Environment (VRE) that indicates the relationship between Research Data Management (RDM) and VREs. The outcome of this study was that VREs are ideal platforms for the management of research data.

In the first part of the study, a literature review was conducted by focusing on four themes: VREs and other concepts related to VREs; VRE components and tools; RDM; and the relationship between VREs and RDM. The first theme included a discussion of definitions of concepts, approaches to VREs, their development, aims, characteristics, similarities and differences of concepts, an overview of the e-Research approaches followed in this study, as well as an overview of concepts used in this study. The second theme consisted of an overview of developments of VREs in four countries (United Kingdom, USA, The Netherlands, and Germany), an indication of the differences and similarities of these programmes, and a discussion on the concept of research lifecycles, as well as VRE components. These components were then matched with possible tools, as well as to research lifecycle stages, which led to the development of a first conceptual VRE framework. The third theme included an overview of the definitions of the concepts ‘data’ and ‘research data’, as well as RDM and related concepts, an investigation of international developments with regards to RDM, an overview of the differences and similarities of approaches followed internationally, and a discussion of RDM developments in South Africa. This was followed by a discussion of the concept ‘research data lifecycles’, their various stages, corresponding processes and the roles various stakeholders can play in each stage. The fourth theme consisted of a discussion of the relationship between research lifecycles and research data lifecycles, a discussion on the role of RDM as a component within a VRE, the management of research data by means of a VRE, as well as the presentation of a possible conceptual model for the management of research data by means of a VRE. This literature review was conducted as a background and basis for this study.

In the second part of the study, the research methodology was outlined. The chosen methodology entailed a non-empirical part consisting of a literature study, and an empirical part consisting of two case studies from a South African University. The two case studies were specifically chosen because each used different methods in conducting research. The one case study used natural science oriented data and laboratory/experimental methods, and the other, human orientated data and survey instruments. The proposed conceptual model derived from the literature study was assessed through these case studies and feedback received was used to modify and/or enhance the conceptual model.

The contribution of this study lies primarily in the presentation of a conceptual VRE model with distinct component layers and generic components, which can be used as technological and collaborative frameworks for the successful management of research data.

Keywords

Virtual Research Environments, VRE, Research Data Management, RDM, core components, pluggable components, component layers, research lifecycle, research data lifecycle, conceptual model

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAF	Australian Access Federation
ACP Secretariat	African, Caribbean, and Pacific Group of States Secretariat
ANDS	Australian National Data Service
ARC	Agricultural Research Council (South Africa)
ARC	African Research Cloud (South Africa)
ARCS	Australian Research Collaboration Services
API	Application Programming Interface
AREN	Australian Research and Education Network, later renamed National Research and Education Network (NREN)
ARL	Association of Research Libraries (USA)
ARMA	Association of Research Managers and Administrators (UK)
ASAUDIT	Association of South African University Directors of Information Technology
ASSAf	Academy of Science of South Africa
CDL	California Digital Library (USA)
CHPC	Centre for High Performance Computing (South Africa)
CIL	Computer Information Literacy
CODATA	Committee On Data for Science And Technology, an interdisciplinary Scientific Committee of the International Council for Science (ICSU)
CPUT	Cape Peninsula University of Technology (South Africa)
CSIR	Council for Scientific and Industrial Research (South Africa)
CVE	Collaborative Virtual Environment
DCC	Digital Curation Centre (UK)
DFG	Deutsche Forschungsgemeinschaft (Germany)
DHET	Department of Higher Education and Training (South Africa)
DIRISA	Data Intensive Research Initiative of South Africa
DMP	Data management plan
DST	Department of Science and Technology (South Africa)
e-IRG	e-Infrastructure Reflection Group (EU)
eRIC	E-Research Infrastructure and Communication project
ESFRI	European Strategy Forum on Research Infrastructures (EU)

ESRC	Economic and Social Research Council (UK)
EU	European Union
HartRAO	Hartebeesthoek Radio Astronomy Observatory (South Africa)
HPC	High Performance Computing
HSRC	Human Sciences Research Council (South Africa)
ICPSR	Inter-university Consortium for Political and Social Research (USA)
ICSU	International Council for Science
IDIA	Inter-University Institute for Data Intensive Astronomy (South Africa)
IFDO	International Federation Of Data Organisations For Social Science
IM	Instant Messaging
JISC	Joint Information Systems Committee (UK)
KNAW	Royal Netherlands Academy of Arts and Sciences (Netherlands)
LERU	League of European Universities (EU)
LIASA	Library and Information Association of South Africa
NCI	National Computational Infrastructure (Australia)
NCRIS	National Collaborative Research Infrastructure Strategy (Australia)
NeCTAR	National eResearch Collaboration Tools and Resources project (Australia)
NeDICC	Network of Data and Information Curation Communities (South Africa)
NICIS	National Integrated Cyber-Infrastructure System (South Africa)
NIPMO	National Intellectual Property Management Office (South Africa)
NREN	National Research and Education Network (Australia)
NRF	National Research Foundation (South Africa)
NSF	National Science Foundation (USA)
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting

OECD	Organisation for Economic Co-operation and Development
OGCE	Open Grid Computing Environment
OSI	Office of Science and Innovation (UK)
PAR	Participatory Action Research
PC	Personal Computer
PDA	Personal Digital Assistant
POL-SABINA	Policy and support actions for Southern African Natural Product Partnership
RCUK	Research Councils UK
R&D	Research and Development
RDA	Research Data Alliance
RDM	Research Data Management
RLUK	Research Libraries UK
SA^3	South African Astroinformatics Alliance
SAAO	South African Astronomical Observatory
SABIF	South African Biodiversity Information Facility
SADA	South African Data Archive
SADC	Southern African Development Community
SAEON	South African Environmental Observation Network
SAGrid	South African National Grid
SALT	Southern African Large Telescope
SAMI	South Africa Malaria Initiative
SANREN	South African National Research Network
SARIMA	Southern African Research and Information Management Association
SARIS	South African Research Information Services Project
SCONUL	Society of College, National and University Libraries (UK)
SIM4RDM	Support Infrastructure Models for Research Data Management (EU)
SKA	Square Kilometre Array
SOA	Service-Orientated Architecture
SPU	Sol Plaatje University (South Africa)
SU	Stellenbosch University (South Africa)
UC3	University of California Curation Center (USA)

UCISA	Universities and Colleges Information Systems Association (UK)
UCT	University of Cape Town (South Africa)
UNISA	University of South Africa
UP	University of Pretoria
VLE	Virtual Learning Environment
VO	Virtual Organisation
VRC	Virtual Research Community
VRE	Virtual Research Environment
WDS	World Data System, an Interdisciplinary body of the International Council for Science (ICSU)
WITS	University of the Witwatersrand (South Africa)
WRSS	Web-based research support systems
XSEDE	Extreme Science and Engineering Discovery Environment (USA)

CHAPTER 1
INTRODUCTION

1.1 CONTEXT OF THE RESEARCH PROBLEM

Virtual Research Environments (VREs) as technology frameworks to facilitate collaborative research projects have been used by a number of universities and research institutions globally. VREs "are an intricate part of e-Research and comprise of digital infrastructure and services (online tools, content, and middleware), which enable research to take place within the virtual multi-disciplinary and multi-organisation partnership context. The specific aim of a VRE is to help researchers manage the increasingly complex range of tasks involved in carrying out research" (Van Deventer, et al., 2009). According to Fraser (2005), VREs arose from the development of e-Science and are intrinsically linked to e-Science. Fraser (2005) also includes cyberinfrastructure and e-infrastructure as part of VREs. He compares them to "Managed Learning Environments (sum of services and systems which together support the learning and teaching processes)" rather than virtual learning environments (VLEs). Fraser regards VREs as "the result of joining together new and existing components to support as much of the research process as appropriate for any given activity or role" (Fraser, 2005). For the purposes of this study, a component is seen as a "uniquely identifiable input, part, piece, assembly or subassembly or subsystem" that is needed to perform "a distinctive and necessary function in the operation of a system" (Business Dictionary, 2018).

The concept of a VRE, according to Voss and Procter (2009: 176), is still somewhat in flux. They point out that the VREs that have been built so far have tended to be either precise configurations for specific research projects, or systems having very generic functions. They also point out that technologies used to build VREs vary extensively, resulting in significant fragmentation and a shortage of interoperability, which necessitates agreed standard platforms and configurable modules that will enable swift development and implementation of tailored VREs accomplishing specific requirements at a reasonable cost, and with moderate effort needed in adoption and adaptation. In other words, there exists a need for the formalisation of a conceptual model of a VRE, that can be used repeatedly in different contexts and different subject fields. This study aims to address that need by focusing on two research projects at a South African university. It should also be noted that research data is recognised internationally as a vital resource of which the value needs to be preserved for future research – assigning a huge responsibility to higher education institutions to ensure that their research data is managed in such a manner that they are protected from substantial reputational, financial and legal risks in the future. In this regard, VREs offer the ideal instruments that could be used in the management of research data.

The purpose of this study is to investigate the relationship between Research Data Management (RDM) and VREs, through the development of a conceptual VRE model that indicates the important role of RDM as a component within a VRE. The purpose of the study will be accomplished through findings from literature, the formalisation of a conceptual model, and by focusing on two case studies at a South African university.

1.2 RESEARCH PROBLEM / QUESTION

How can a Virtual Research Environment be conceptualised to indicate the role of Research Data Management (RDM) within a VRE?

In order to address the stated research problem, the following sub-questions will be asked:

- What is a VRE?
- What is the current state of VRE research in the world?
- What are the generic components that make up a VRE?
- How does a VRE support a research cycle?
- What is RDM?
- Why should a VRE be an essential technological and collaborative framework for the management of research data?
- To what extent can the components identified through the third sub-question be formalised into a conceptual model?
- Where would RDM as a component be placed?
- To what extent can this model be generalised for use in other environments?

1.3 RELEVANCE OF STUDY FOR THE SUBJECT FIELD

The study aspires to contribute to the subject field in the following manner:

- Understanding how VREs operate;
- Identifying generic components of VREs;
- Understanding the concept of RDM, related concepts, and their characteristics;
- Comprehension of research data lifecycles;
- Identifying generic research data lifecycle components, as well as actions in each of the components;
- Understanding the essential role that the management of research data plays in VREs, and how VREs are used to manage data; and
- Developing a possible conceptual model that can be applied in other disciplines and multidisciplinary settings.

1.4 RESEARCH METHODOLOGY

1.4.1 Research Design

In this section, a brief overview is provided of the research design followed in this study (for a more detailed overview, see Chapter 6).

This study follows an interpretivist paradigm, with a focus on empirical interpretivism. Empirical interpretivism focuses on the investigation of social phenomena in natural settings (Pickard, 2007: 11), which is ideal when studying a VRE as a 'social phenomenon' in its natural setting.

The approach followed in this study is qualitative in nature. Qualitative research is described by Creswell (2007: 37) as "the possible use of a theoretical lens, and the study of research problems inquiring into the meaning individuals or groups ascribe to a social or human problem."

The research design used in this study consists of both empirical and non-empirical study. The non-empirical part of this study includes a literature review, and the empirical part is focused on two VRE case studies at a South African University. Both these case studies were closed communities in the sense that they each consisted of a supervisor and researcher students, and were concentrated on a specific research area; however, researchers from other communities could join these groups or could request access to these case studies.

1.4.2 Literature Review

The aim of the literature review was to help define the key concepts and to lay a theoretical framework for the empirical study.

The literature review comprises the following chapters:

- Chapter 2 – VREs as part of e-Research infrastructure and other related concepts;
- Chapter 3 – VREs and their components;
- Chapter 4 – Research Data Management (RDM);
- Chapter 5 – Relationship between RDM and VRE. (It should also be noted that this chapter includes a conceptual VRE model, which was developed from information that was gained from the literature review. This conceptual model was then tested in the empirical part of the study).

1.4.3 Case Studies

The empirical part of the study, which focuses on two case studies, is covered in Chapter 7. In this chapter, the results gained through a formative evaluation process, are discussed, followed by a discussion of results gained from a summative evaluation process through interview questions. Data from these case studies were collected through a method of Participatory Action Research (PAR) in conjunction with prototyping, by using the following data collection tools: observation, semi-structured interviews, and testing/experimenting.

Purposive sampling was chosen to identify respondents, because of the researcher's knowledge of the researchers involved, as well as their roles and characteristics within the VREs. Criteria were identified from the conceptual framework model (which was compiled through the literature review), and used to compile a broad profile of the categories of participants needed, who could be approached to give insight into the research problem. The whole population of these case studies were selected in this process.

1.4.4 Methods Of Analysis

Methods used to analyse the results gained through the observations, the semi-structured interviews and testing/experimenting, included pattern-matching, explanation building, and analysis of tools.

1.5 LIMITATIONS OF THE STUDY

The following limitations of the study should be taken into consideration:

- The study only focused only on two case studies within one institution, which inhibited the possibility to generalise the proposed conceptual framework model;
- The research was only conducted within an academic context, with individual researchers / students conducting research with the aim of gaining a degree, and where face-to-face contact was relatively easy, and not with researchers from a variety of different research organisations working from remote locations on the same project; and
- The study only focused on small data sets, and not on big data.

1.6 STRUCTURE OF THE THESIS

The thesis is structured as follows:

Chapter 1	Introduction	
Chapter 2	VREs as part of e-Research infrastructure and other related concepts	Literature Review
Chapter 3	VREs and their components	
Chapter 4	Research Data Management (RDM)	
Chapter 5	Relationship between RDM and VRE	
Chapter 6	Research Methodology	
Chapter 7	Results: Presentation and Discussion	Empirical Study
Chapter 8	Concusions and Recommendations	

1.7 EXPOSITION OF CHAPTERS

- **Chapter 1: General Introduction**

  This chapter sketches the context to the research problem/question, and lists the research problem/question and its sub-questions. The relevance of the study to the research field is indicated, followed by a brief overview of the research methodology followed. Finally, an exposition of the chapters is discussed.

- **Chapter 2: Virtual Research Environments (VREs) as part of e-research infrastructure, and other concepts related to VREs**

  In this chapter, a background to VREs is provided. This is done by discussing related concepts to VREs, such as e-Research, e-Science and cyberscience, cyberinfrastructure, collaboratories, science gateways and Web-based Research Support Systems (WRSS). The different approaches to VREs are then examined, followed by an overview of the concept 'Virtual Research Environments', its definition, development, aims and characteristics. In conclusion, the researcher highlights the similarities and differences of above-mentioned concepts with VREs, and gives an indication of the e-Research approaches this study followed, as well as the concepts used in this study.

- **Chapter 3: VREs and their components and tools**

  This chapter commences with an overview of the developments of VREs across the world, specifically focussing on four countries (United Kingdom, the USA, the Netherlands, and Germany), followed by an overview of the differences and similarities in the programmes of these countries. The discussion then turns to the concept of research cycles as well as the various VRE components as found in the literature. Thereafter, the research components as well as possible VRE tools are matched to the research cycle stages. Finally, the researcher presents a possible conceptual framework that combines insights gained from VRE developments internationally, the research cycle concept, the VRE components, as well as VRE tools.

- **Chapter 4: Research Data Management (RDM**)
  This chapter starts with an overview of the concepts data and research data, as well as related concepts, and how each relates to RDM. The discussion also includes the concepts data curation, data stewardship, data governance, data archiving, and data management. RDM is then defined, followed by an overview of a number of international developments with regards to RDM. Thereafter, the different approaches to RDM are compared and the similarities and differences highlighted. The South African situation with regard to RDM is examined next, discussing government initiatives, national collaborative initiatives, initiatives at higher education institutions, other initiatives and potential partners. After this, the concept of a research data lifecycle is explored by comparing a number of cycles from literature, followed by a discussion of the different stages of a research cycle as well as the corresponding processes that takes place in each, and the potential role that the various stakeholders can play in each. The chapter is concluded with a discussion on the processes that takes place throughout the whole lifecycle.

- **Chapter 5: RDM and VREs**
  The aim of this chapter is to explore the relationship between RDM and VREs. This is done by first looking at the research data lifecycle and its relation to the research lifecycle. Next, the role of RDM as a component within a VRE is explored, followed by a discussion of the management of research data by means of a VRE. Following this a possible conceptual model for the management of research data by means of a VRE is presented. The chapter is concluded by a short overview / synopsis of the literature review.

- **Chapter 6: Research Methodology**
  In this chapter, an outline is given of the research methodology that was followed in this study. An overview is presented of the non-empirical part of the study, which comprises a literature study of the concepts, as well as the empirical part of the study, which constitutes case studies. A description is also given of what is meant by a literature study, and by a ‘case study’ method. Various methods used in the case study are also discussed, namely sampling method, triangulation, PAR, and prototyping. The discussion further focuses on the various data collection methods used, namely participant observation, interviews, as well as testing and prototyping. Finally, an overview is given of the research questions asked during the interviews, followed by a description of the methods of analysis and of evaluation.

- **Chapter 7: Results: Presentation and Discussion**
  The analysis of findings from two case studies is presented in this chapter through a process of formative and summative evaluation. The formative evaluation was done through a process of PAR using the following data collection techniques: notes taken during meetings, training sessions with the members of these case studies, as well as e-mail correspondence between the VRE design team and members of these VRE groups (Case Study A and Case Study B). A process of testing and prototyping were used to identify and design suitable platforms (tools) for a technological framework for a VRE, which could be used in these case studies. The formative evaluation was then followed up by a process of summative evaluation, consisting of semi-structured interviews with the members in each of these case studies. The answers received were then mapped to findings in literature as well as well as to results received through the formative evaluation process.

- **Chapter 8: Conclusions and Recommendations**
  The aim of this chapter is to address the central research question and its sub-questions from the findings in the empirical part of the study, corroborate these from findings in the literature study, and to draw conclusions from these. This is followed by a reflection on the findings that were gained from the case studies and the literature. Next follows a discussion about the contribution of this study to the subject field, and an indication of the limitations of the study. The study is concluded by the researcher listing a number of guidelines and recommendations for setting up a

conceptual VRE model, presenting suggestions for further research, and providing concluding remarks.

1.8 SUMMARY

This chapter sketched the context of the research problem and listed the research problem/question as well as its related sub-questions. Next, the relevance of the study for the subject-field was highlighted. After this, a brief overview was given of the research methodology followed in this study. The limitations of the study were then highlighted and lastly, an exposition was given of the structure of the eight chapters.

In the next chapter, VREs as part of e-Research infrastructure and other concepts related to VREs are discussed.

CHAPTER 2
VREs AS PART OF E-RESEARCH INFRASTRUCTURE, AND OTHER CONCEPTS RELATED TO VREs

2.1 INTRODUCTION

Since the introduction and development of advanced information technologies, research practices across a wide range of disciplines have changed profoundly. These new and emergent changes were first recognised in the science and engineering disciplines, and were first described using terms such as e-Science or cyberscience (Lynch, 2008: 74). These concepts have evolved over the years and broadened into the concept of e-Research, which include the social sciences and arts and humanities (Allan, 2009: 5). This has led to the development of VREs or similar terms such as cyberinfrastructure or collaboratories, in order to provide the necessary technology frameworks in support of e-Research.

To gain clear understanding of the concept of VREs, and later of RDM, the researcher conducted a literature review by exploring books, reports, guides, journal articles, and blogs that were written on the topics, as well as web sites of institutions involved with VRE initiatives and RDM initiatives.

The aim of this chapter is to first sketch a background to VREs by discussing related concepts such as e-Research, e-Science and cyberscience, cyberinfrastructure, collaboratories, science gateways and WRSS. The different approaches to e-Research are also discussed. This is then followed by an overview of the concept 'Virtual Research Environments' - its definition, development, aims, and characteristics. Finally, the similarities and differences of the above-mentioned concepts with VREs are highlighted, together with an indication of the e-Research approaches this study followed, and an indication of the concepts to be used specifically in this study.

2.2 KEY CONCEPTS

In the discussion of the concept of VREs, it is imperative to first look at closely related concepts as found in the literature.

2.2.1 E-Science

The term e-Science was coined in 1999 in the United Kingdom (UK) by John Taylor, Director General of the Research Councils in the UK Office of Science and Technology (OST) from 1999-2003. It is described as "global collaboration in key areas of science and the next generation of infrastructure that will enable it" (Hey and Trefethen, 2003: 1017; Jankowski, 2007: 551). The idea of using technology infrastructure to support collaborative research was already supported by earlier researchers, though, for example Kraut, Egido and Galegher (1988: 1), who studied research collaboration and the technological needs of these collaborative groups. They describe science as a "fundamentally social process" where "the development of new ideas for scientific research, the execution of research tasks, and the preparation of formal research reports are all processes that involve extensive social interaction." Kraut, Egido and Galegher (1988: 1) were of the view that an understanding of the nature of collaborative work relationships would be helpful in the testing and implementation of technologies that support collaborative research. Hey and Trefethen (2008: 15) also include this aspect of collaboration in their definition of e-Science as "the next generation of scientific problems, and the collaborative tools and technologies that will be required to solve them."

The idea of collaboration can also be found in Hwang, Capellan and Consulta's (2008: 63) description of e-Science as "a new approach to science involving distributed global and international collaborations enabled by the Internet and using very large data collections, terascale computing resources and high-performance visualizations." E-Science according to them is about "global collaboration in key areas of science, and the next generation of infrastructure, namely the grid, that will enable it" (Hwang, Capellan and Consulta's, 2008: 63). Their idea of 'global collaboration' is affirmed by Dutton and Jeffreys (2010: 6), who associate e-Science with worldwide collaboration in key areas of science" and being involved in "the building of e-infrastructures that can sustain the long-term growth of e-research." Dutton and Jeffreys (2010: 6) furthermore point out that the term e-Science is used by the UK's e-Science Programme, which describes e-Science as "the systematic development of research methods that exploit advanced computational thinking."

Hey and Hey (2006: 516) emphasize the idea of networks and vast amounts of data as well as the development of infrastructure when they describe e-Science as "a new research methodology" made possible by "networked and data-driven science." E-Science, according to ARL (2007: 6), "requires new strategies for research support and significant development of infrastructure." Münch (2011: 30), in turn, regards the term e-Science as a broader term than just electronic science. She refers to it as "enhanced science." According to Münch (2011: 30), "it describes an integrated digital infrastructure for scientific publication, collaboration, and information exchange."

Wouters and Beaulieu (2006: 62) point out that e-Science in the UK originally had been about constructing a specific type of infrastructure for research, but that the infrastructure was developed in the context of a very specific epistemic culture originating from high energy physics, which was subsequently adapted for computer science and bioinformatics contexts. E-science in this context can therefore generally be defined as "the sharing of computational resources, [high performance computing], distributed access to massive datasets, and the use of digital platforms for collaboration and communication" (Beaulieu and Wouters, 2009: 55, 56). In order to expand the concept of e-Science to cover more subject areas, the UK established a government-sponsored office to encourage and coordinate e-Science in the social sciences in 2004 (Jankowski and Caldas, 2004). This office, called the National Centre for e-Social Science (NCeSS), included a decentralised structure of nodes that connected universities across the UK (Jankowski, 2009: 5). The principal aim of NCeSS was to enable social scientists to utilise innovations in digital infrastructure, in order to be equipped in addressing crucial challenges in their respective subject fields (Halfpenny et al., 2009: 73).

Beaulieu and Wouters (2009: 56) argue that although e-Science was expanded to cover social science under the banner of e-Social Science, it did not cover non-computational e-Science and research not reliant on high performance computing (HPC). They argue for a more inclusive concept, namely ‘e-Research’, which would cover all aspects of e-Science, including non-computational e-Science, research not reliant on HPC, as well as other methods of using new media and digital networks, for example e-mail, websites, social media, etc. (Beaulieu and Wouters, 2009: 56). The researcher decided to follow the approach of Beaulieu and Wouters (2009: 56) and to approach this study from an e-Research perspective. This is particularly important in the empirical part of this study, when focusing on VRE pilot studies from different subject areas.

2.2.2 Cyberinfrastructure

A term sometimes used synonymously with e-Science is cyberinfrastructure. The concept has its origin in initiatives in the natural and biological sciences in the USA and is basically an American version of the European concept of e-Science. The term ‘cyberinfrastructure’ was originally advocated in a commissioned report financed by the USA’s National Science Foundation (NSF) in 2003, titled ‘Revolutionizing Science and Engineering Through Cyberinfrastructure’, which in due course became known as the Atkins Report (2003) (Jankowski, 2009: 5). According to this report, cyberinfrastructure refers to infrastructure “based upon distributed computer, information and communication technology.” Furthermore, “if infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy” (Atkins et al., 2003: 5). Indiana University provides a description of cyberinfrastructure in a similar vein to Jankowski (2009: 6), but more detailed, as consisting of “computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high-performance networks to improve research productivity and enable breakthroughs not otherwise possible” (Stewart, 2010).

Jankowski (2009: 6) further argues that initial cyberinfrastructure initiatives were in the natural and biological sciences, which dealt with research consisting of large quantities of data and which require high-speed computer processing, for example astronomy, meteorology, particle physics and DNA research. The emphasis on cyberinfrastructure has also been found in the areas of science and engineering, in which large quantities of data are involved and are processed with the assistance of “grid computer networks and related software” (Jankowski, 2009: 6). Cyberinfrastructure, however, is not only applicable to science, engineering and biological sciences; it can also be applied to other disciplines, as is clear from the Atkins Report (2003) (Jankowski, 2009: 5). Although geared towards science and engineering, the report specifically states that the scope of cyberinfrastructure extends to all research fields and education (Atkins et al., 2003, 31). The NSF subsequently held a workshop on cyberinfrastructure and the social sciences in 2005, and produced a report describing how cyberinfrastructure can facilitate social science research and showing that the social sciences and behavioural sciences can contribute significantly to the advancement of cyberinfrastructure (Berman and Brady, 2005: 4, 6). The report on cyberinfrastructure for the humanities and social sciences, which was issued by the American Council of Learned Societies (ACLS) in 2006, further enhanced the idea of cyberinfrastructure for the humanities and social sciences (ACLS, 2006). This report describes and gives an analysis of the state of humanities and social science cyberinfrastructure, and outlines what the requirements and potential contributions of the humanities and social science are in constructing a cyberinfrastructure for information, research and teaching (ACLS, 2006: 1).

Cyberinfrastructure, furthermore, is described by Unsworth (2006: 6) as “more than a tangible network and means of storage in digitized form, and it is not only discipline-specific software applications. It is also the more intangible layer of expertise and the best practices, standards, tools, collections and collaborative environments that can be broadly shared across communities of enquiry.”

A concept closely related to cyberinfrastructure is science gateways.

2.2.3 Science Gateways

The science gateway concept has been widely used in the USA, and can best be described as a community-developed bundle of tools, applications, and data collections customised to meet the needs of a targeted community, which are integrated via a portal, or a collection of applications (Indiana University, 2012; Wilkins-Diehr, 2007: 743; Yang and Allan, 2010: 69). A science gateway provides an interface between a researcher (or community) and distributed computing infrastructures (LPDS, 2013). This includes access to and enabling of interoperability between e-infrastructures consisting of different middleware and architectures, for example grid, HPC, cloud or simply local clusters (Chain-Reds, n.d.). These gateways provide entrance to an array of "capabilities including workflows, visualization as well as resource discovery and job execution services" (Wilkins-Diehr, 2007: 743; Indiana University, 2012; Wu et al., 2010). Science gateways also provide collaborative cyber-environments, enabling researchers working in similar domains to team up easily to perform "computational thinking" and research regarding "challenging scientific problems" (Wu et al., 2010). In addition, science gateways provide a means for users to store, manage, catalog, and share large data collections, or speedily develop novel applications they cannot locate elsewhere (Wilkins-Diehr et al., 2008: 33).

According to Allan (2009: 15) and Indiana University (2012), science gateways can take one of three formats:

- A gateway, packaged as a web portal with front-end users accessing dynamic distributed computing infrastructures behind it, for example, grid services, cloud computing, and high-performance computing;
- A grid-bridging gateway, enabling communities to run their own grids devoted to their areas of science, and by doing so, extending the reach of the community's grid by coupling it with distributed computing infrastructures; and
- A gateway that allows users to run rich desktop applications that can access distributed computing infrastructures.

Basney et al. (2011) emphasize that science gateways cannot be described as cyberinfrastructure themselves, but can offer convenient interfaces to cyberinfrastructure and can broaden and simplify its use, without the need to understand all of its intricacies.

Another term used synonymously with e-Science is cyberscience.

2.2.4 Cyberscience

Cyberscience is a more recent term, used by Nentwich (2003: 3). He regards the term as synonymous with e-Science. He defines cyberscience as "all scholarly and scientific research activities in the virtual space generated by the networked computers and by advanced information and communication technologies in general." The term, according to Nentwich (2003: 22, note 41), originated from an article by Paul Wouters titled 'Cyberscience', which appeared in the Dutch journal *Kennis en Methode* in 1996. Following this article, the term appeared in various papers and conference panels, but use of the term has since mainly been limited to publications and projects emerging from the Austrian Institute of Technology Assessment, Nentwich's institutional home, and a recent study by Christine Hine (2008), 'Systematics as Cyberscience' (Jankowski, 2009: 3). The core feature present in both Wouters' and Nentwich's formulation of cyberscience is: "an all-encompassing approach that acknowledges the importance of computers and electronic networks, but that is grounded in a broad vision of the scholarly enterprise" (Jankowski, 2009: 3). This approach includes scholarly communication and publishing in all disciplinary areas, linking up with the formulation of another concept, e-Research, which is discussed next.

2.2.5 E-Research

The concept of e-Research has been discussed by various authors, each with their own unique definition or description of the term.

E-Research according to O'Brien (2005: 66) is a broader term than e-Science in that it includes non-scientific research, but also "refers to large-scale, distributed, national, or global collaboration in research," and, according to the Australian Research Council (2005: 2), typically "entails harnessing the capacity of information and communication technology (ICT) systems, particularly the power of high capacity distributed computing, and the vast distributed storage capacity fuelled by the reducing cost of memory, to study complex problems across the research landscape." Lawson and Butson (2007: 2) confirms O'Brien's (2005: 66) viewpoint by describing e-Research as a blanket term that "covers the entire general area of information and communication technologies (ICTs)" that are supporting researchers in their research cycle (Lawson and Butson, 2007: 2). They regard e-Research as an extension of the e-Science concept. Beaulieu and Wouters (2009: 55), in turn, distinguish between e-Science and e-Research. According to them, there are differences in the integration of cyberinfrastructures and tools. The concept of e-Science, according to Wouters and Beaulieu (2006: 56) and Beaulieu and Wouters (2009: 55), emphasises "data-orientated, computational or quantitative analysis", whereas e-Research covers forms of non-computational e-Science, in other words research not reliant on HPC, as well as other methods of utilising new media and digital networks (Beaulieu and Wouters, 2009: 56).

Borgman (2007: 20) and Jankowski (2009: 6) describes e-Research as more reflective of initiatives in the social sciences and humanities, which is in contrast to the discussion in 2.2.1, which showed that later initiatives in e-Science also included the social sciences and humanities. Borgman (2007: 20) and Jankowski (2009: 6) nevertheless elaborate on this difference. E-Research as a concept, according to them, "acknowledges forms of scholarship that do not primarily emphasize use of high-speed computers for processing large datasets, but place[s] [sic] weight on incorporation of a wide variety of new media and electronic networks in the research process;" aspects normally found in e-Science (Borgman, 2007: 20; Jankowski, 2009: 6). Jankowski (2009: 6) further describes e-Research as "a form of scholarship conducted in a network environment utilizing Internet-based tools and involving collaboration among scholars separated by distance, often on a global scale." Dutton and Jeffreys (2010: 6) also emphasize the inter-disciplinarity of e-Research and the wider scope than the use of high speed computers for processing large datasets in their definition. E-Research, according to them, "refers to the wide array of research activities" and "is not tied to a particular disciplinary area, and is therefore suited" to include "the broader potential of the application of ICTs (Information Computer Technologies) to research."

For the purpose of this study, the aspects in the abovementioned definitions can be integrated in the following definition of e-Research:

> E-Research can be defined as a broad term that extends e-Science. It is a form of scholarship conducted in a networked environment that includes all ICTs that support researchers in their research process. This includes all forms of non-computational e-Science, consisting of a wide variety of new technologies, tools and computer networks, which can be used collaboratively by researchers and that can be co-located or separated by distance globally.

Various approaches to e-Research can be found in the literature.

2.2.5.1 Approaches to e-Research

Fry and Schroeder (2009) as well as Searight et al. (2011: 71) distinguish between two main types of approaches to e-Research: the conventional approach and the social science approach.

(a) Conventional approach to e-Research

In this approach, the focus, according to Fry and Schroeder (2009: 37), is on research apparatuses (technology) for manipulating data and the physical environment. This approach is dominated by computer science and its development of new technologies. Searight et al. (2011: 71) regard conventional e-Research as synonymous to e-Science, describing it as relating "to applications of grid architecture for accessing and analysing large data sets from a geographic distance." This approach, according to Hey and Trefethen (2003: 809, 811), concerns itself with data and the ability to cope with the "data deluge" (the huge expansion in the volume of data expected from "the next generation of experiments, simulations, sensors, and satellites, etc." that need to be curated and stored and made accessible in different ways). This data deluge will have a profound influence on existing scientific infrastructure. Data will be generated from a variety of new sources, and "will need to be annotated with metadata, archived and curated" so that it can be searched and analysed, reproduced and visualised in the future (Hey and Trefethen, 2003: 821).

Kahn (2004) focuses on another characteristic of conventional e-Research, namely the use of computational science. Computational science comprises the appropriate utilisation of computational architecture (e.g. ICTs and scientific calculators) in the application of an algorithm or method (e.g. information processing, simulation and modelling of complex phenomena) to solve some scientific application or method of real-world scientific or societal interest (CSERD, 2012; Jeffreys, 2010: 51; The College at Brockport, 2012). Kahn (2004) further predicts that computational science would solve more complex research problems in the near future with greater accuracy, going deeper, being applied by more scientists, more routinely and going wider. Its greatest impact, however, would be in its role of breaking down barriers between 'silos' of scientific domains, and enabling real e-Research to take place (Kahn, 2004).

(b) Social Science approaches to e-Research

E-Research, according to Fry and Schroeder (2009: 35-53), can also be approached from a social science perspective. They emphasize that e-Research has generally increased the scope and scale of collaboration and communication, resulting in tremendous organisational problems, but also extending the technological infrastructure of research. In other words, “e-Research faces new challenges involving [the] ‘control’ and ‘coordination’ of research” (Fry and Schroeder, 2009: 37). On the one hand, there are research instruments for manipulating the objects of research, and on the other hand, there are instruments for communication and collaboration, but the balance between the two, according to Fry and Schroeder (2009: 37), is still unclear. Fry and Schroeder (2009: 38-39) further divide their social science approaches in two categories: those studies concerned with e-Research as an object of research, and those studies concerned with e-Research as an object of development. They base these two divisions on two dimensions:

- The degree to which approaches are pragmatic (where the focus is on practical aspects of development and use of e-Research tools and resources); and
- The degree to which approaches attempt to engage with research.

These two dimensions can be on a pro-active level or can take a detached stance. Fry and Schroeder (2009: 38-39) plot the approaches they have identified along these two dimensions (conceived as continuous, rather than fixed) resulting “in a taxonomy consisting of four main categories: proactive-engagement/pragmatic; proactive-engagement/research; detachment/pragmatic and detachment/research.” This taxonomy then delivers eight approaches (Fry and Schroeder, 2009: 39), as illustrated in Table 2.1. Though the approaches are numbered, they are in no specific order.

The researcher of this thesis adapted the two tables from Fry and Schroeder (2009: 39) by combining them into one table (Table 2.1), in order to illustrate the different approaches more clearly. The headings of the two tables in Fry and Schroeder (2009: 39), namely Approaches to e-Research as Object of Research and as an Object of Development were placed in an added column at the right side of Table 2.1.

Table 2.1: Adaptation of Fry and Schroeder's (2009: 39) approaches to e-Research

	Pragmatic	***Research***	
Pro-active Engagement	**Approach 1: Usability/practical** "e.g. How appropriation can be enhanced through refining understanding of practice, user representations, and human computer interaction."	**Approach 2: Value free/attempted neutrality** "e.g. Measuring dimensions of distributed communication and collaboration."	**Approach to e-Research as an Object of Research**
	Approach 3: Agenda Neutral/Supporting Paradigms "e.g. Concern with tools being user-led; development efforts addressing user-needs in a specific research paradigm, for example discourse analysis, gene ontologies, text-corpora in linguistics. Social factors perceived as technology and policy re-engineering problem."	**Approach 4: Embedded in the Disciplines / Sustainability in Adoption** "e.g. Emerging from a positivistic tradition. Addressing computational and processing issues for domain-specific problems; uptake and use perceived as a result of overcoming technical problems."	**Approach to e-Research as an Object of Development**
Detachment	**Approach 5: Advocacy/steering and aligning structures** "e.g. Fostering institutional, economic, and legal structures that enable distributed communication and collaboration. Promoting a particular type of open and accessible e-research."	**Approach 6: Critique/reflexive or prospective** "e.g. Social implications of e-research; ability to deliver on claims; policy."	**Approach to e-Research as an Object of Research**
	Approach 7: Agenda Aligned /Supporting Generic Infrastructure "e.g. Concern with development of services across the disciplinary boundaries; social factors perceived as a social re-engineering problem."	**Approach 8: Scepticism / Non-use from within the Disciplines / Sustainability as Project** "e.g. Possibly leading to resistance; uptake and use related to perceived relevance of e-research."	**Approach to e-Research as an Object of Development**

The following approaches to e-Research can also be classified as social science approaches:

Computerisation Movement Approach

De la Flor and Meyer (2008: 1) also approach e-Research from a social science perspective by framing e-Research as a ‘computerisation movement’ (CM) using Iacono and Kling’s (2001) conceptual model. According to Iacono and Kling (2001: 94), “the meaning of the Internet is being built up or ‘framed’ in macro-level discourses such as those of the government, the media and scientific disciplines.” They conceptualise the processes of spreading these frames across many layers of public discourse and the resultant mobilization of large-scale support and specific lines of action within micro-social contexts (e.g. the restructuring of organisations so that internetworking technologies can be implemented and used effectively in their routine activities) as CMs (Iacono and Kling, 2001: 94). A CM is described by Kling and Iacono (1988: 228) as a “kind of movement whose advocates focus on computer-based systems as instruments to bring about a new social order." CMs, according to them, encourage and develop ideological beliefs about "what computing is good for" and how partakers in these projects "should manage and organize access to computing" (Kling and Iacono, 1988: 227). Their main thesis is that "computerization movements communicate key ideological beliefs about the links between computerization and a preferred social order” supporting the legitimisation of computerisation for many potential adopters (Kling and Iacono, 1988: 227).

Iacono and Kling (2001: 99) in their study of CMs also focus on a process of societal mobilisation with three primary elements that are related to one another: technological action frames, public discourses, and organisational practices. Technological action frames are a core set of understandings about what a technology is and how it works, and how it is envisaged to look in the future (Hine, 2006: 30; Iacono and Kling, 2011: 99). The technological action frames are made available through public discourses (oral and written public communications), while also shaping and structuring these discourses. These public discourses are then applied at local level through individuals and organisations, and shape and structure organisational practices (Hine, 2006: 30; Iacono and Kling, 2001: 30).

De la Flor and Meyer (2008: 3) explore the possibility of attempts made by various e-Research initiatives to communicate a set of ideological beliefs that could link the use of specific technologies with a preferred approach to research. They also look into the reasons for this. The first study they refer to is a study by Hine (2006). Hine's (2006: 1, 27) study about CMs and scientific disciplines focuses on science and technology studies and its application to e-Science. Even though her study focuses on e-Science 'per se', the results of the study could potentially also be applied to e-Research as a wider concept. Hine (2006: 29) points out that there is "an overall move towards the development of new information infrastructures for science and notes that these technologies may be utilised in various different ways across scientific institutions, research fields, and disciplines, depending on how they are introduced. To understand the uses of these technologies, Hine (2006: 29) stresses the importance of "examining the process by which new infrastructures for science are developed." The introduction of new technologies very often could carry biases, beliefs and symbolic qualities, for example the belief "that computers can be a source of societal transformation." Hine (2006: 29) then applies the CM as a framework to explore these beliefs about computing.

Hara and Rosenbaum (2008: 234) also classify e-Science as a CM, describing it as "a fairly constrained CM that encourages discourses among a specialized population," and although its scope is small, they see it as "a CM because technological action frames have arisen around it," giving rise to various types of discourses and involving a variety of computerisation practices within a string of organisations. Although Hara and Rosenbaum (2008: 234) use the term e-Science, this could, as in Hine's (2006) study, also potentially be applied to e-Research as the wider concept. Hara and Rosenbaum (2008: 232) approach the concept of CMs from another angle. They show in their study that some CMs can be enacted outside of organisations, and that CM's are not bound by organisational structures. They also group Kling and Iacono's classification of general CMs into "finer grained categories", which means that it is much more complex to develop a single set of criteria when analysing the success or failure of a CM (Hara and Rosenbaum, 2008: 232). They identify five criteria pairs (forming continua) within which different CM's can be plotted: external – internal, market-driven – non-market-driven, wide – narrow, stand-alone – bundled, and positive – negative (Hara and Rosenbaum, 2008: 233-234). They code e-Science as fully internal, narrow, in the middle between market-driven and non-market-driven, fully bundled, and positive (Hara and Rosenbaum, 2008: 236).

Information Systems Approach

Information systems (IS) is described by Avgerou (2000: 567) as an academic field that originated in the applied computer science studies field and originally "aimed at systematising the design of data processing applications in organisations." The focus of most information system studies, according to him, is the human organisation. The IS field has subsequently broadened its scope to include an investigation into the attempts organisations make to respond to the challenge of continuous innovation in ICT. This still-expanding scope also includes a focus on the wider context within which an organisation is embedded (Avgerou, 2000: 568). He then identifies five thematic areas of IS research: "applications of ICT to support the functioning of an organisation; the process of systems development; information systems management; the organisational value of information systems and the societal impact of information systems" (Avgerou, 2000: 568-569). McDonald (2005: 145-146), in turn, describes IS as "an active interventionist discipline" that focuses on the "interaction of information technologies with human activity systems." IS, according to him, mobilises information and knowledge so that one is able to "take knowledgeable and informed actions" in his/her "social and organisational setting." It also helps in clarifying and formalising areas of human activity and creating "IT-based systems that can intervene in those areas" to profit all (McDonald, 2005: 145-146).

McDonald (2005: 147) names research as an example of a human activity system, which is then analysed at different levels of granularity:

- **Personal level**: At the personal level, research involves issues of knowledge, motivation, skill and personality, which the researcher brings to his/her work. Activities at the personal level include literature studies, research design, data collection, analysis and the publishing of research. Technologies used at the personal level include document management systems, data collection and management systems, and analytic tools (McDonald, 2005: 147).

- **Social level**: At the social level, research “consists of the personal networks, public behaviours, norms and culture that are exhibited by research groups and collaborations.” Technologies used at the social level include communication technologies, e-mail, video-conferencing tools, collaborative tools, and social networking tools (McDonald, 2005: 147).

- **Organisational level**: Research at the organisational level, according to McDonald (2005: 147), includes “the processes, accountability and power structures in organisations,” for example universities, scientific and research institutes, the military and parts of industry. The ICT infrastructure is owned by the organisation at this level.

- **Societal level:** At the societal level, research concerns itself with questions “about who pays for, and who benefits from research” (McDonald, 2005: 148).

McDonald (2005: 149) further suggests aspects that an IS approach to e-Research could include:

- “Research data warehouses;
- Ontological systems for content organisation;
- Meta-analysis to accumulate work with a similar ontological basis;
- More advanced techniques of domain analysis;
- Knowledge management mechanisms for evidence-based research;
- Serious e-libraries; and
- Development of domain specific patterns.”

McDonald (2005: 150) concludes that ICT does not work effectively in human activity systems, e.g. research, if IS is excluded.

Service Oriented Architecture (SOA) Approach

A survey of literature shows that a great variety of definitions of service oriented architecture (SOA) exist. Some of these will be briefly discussed.

Schroth and Janner (2007: 36) describe SOA as “the philosophy of encapsulating application logic in services with a uniformly defined interface, and making these available via discovery mechanisms.” The Oasis Reference Model for Service Oriented Architecture 1.0 (2006), in turn, defines SOA as a paradigm where distributed capabilities that might be under the jurisdiction of diverse ownership domains, are organised and utilised. It also presents a “uniform means to offer, discover and use capabilities” so that “desired effects consistent with measurable preconditions and expectations” are produced. A service provider might list a well-defined interface on a registry so that other stakeholders are able to “retrieve and loosely couple the offered service with their own services” (Schroth and Janner, 2007: 37). Valipour et al. (2009: 34) describe SOA in terms of business processes and define it as an “architecture that modularizes services.” These services can also be recombined “in various forms for the implementation of new or improved business processes.” They furthermore describe SOA as a “design that links business and computational resources (e.g. organisations, applications and data) on demand to achieve desired results” (Valipour et al., 2009: 35).

According to Sim et al. (2005: 2), a SOA is “a style of design that guides all aspects of creating and using services through their lifecycle (from conception to retirement), as well as defining and providing the information infrastructure that allows different applications to exchange data regardless of the operating systems or programming languages underlying those applications.” These services are described by Sim et al. (2005: 2) as being “coarse-grained, reusable IT assets that have well-defined interfaces” with a clear separation between the services’ externally accessible interface and its technical implementation. This separation decouples or disconnects requesters from service providers, making it possible for both to develop individually “as long as the interfaces remain unchanged.” This characteristic of loose coupling makes it possible for researchers to develop and deploy applications incrementally (Makola et al., 2006). Added to this is the ability to attach new features easily after the system is deployed, and the ability to re-implement existing features to take advantage of developments in hardware or software. This modularity and extensibility according to Makola et al. (2006), is what make SOA especially suitable as an approach for the collaborative research (e-Research) environment.

A SOA approach to VREs ensures that they are flexible enough for dynamically changing user needs. Such an SOA approach to VREs makes it possible to compile a bundle of services that meet the stated tacit user needs, for example, authorisation, authentication, and communication services. These core services can be expanded (by plugging in new services, or making use of external services) to meet new needs (Yang and Allan, 2006a: 454).

Whole Process Approach

According to Paterson et al. (2007: 128). the 'whole process approach' to 'Revolutionizing Science and Engineering Through Cyberinfrastructure' comprises information processing, information exchange, information utilisation and information management. This approach is also involved in the development of demonstrator models to illustrate how the process will work in practice (Paterson et al., 2007: 128). Additionally, this approach grants an in-depth examination of the technical and information management issues involved in e-Research (including the comprehensive issues of establishing an appropriate security framework). An opportunity to examine new legal and policy issues is also afforded.

This study follows the social sciences approach to e-Research, including computerisation movement, IS, SOA and whole process approaches.

Another term which is closely related to VREs, are 'collaboratories'.

2.2.6 Collaboratories

The concept collaboratory, which is a "hybrid of collaborate and laboratory" (Carusi and Reimer, 2010: 14) was first coined by William Wulf in 1989. He described a "center without walls, in which the nation's (USA) researchers can perform their research without regard to physical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, [and] accessing information in digital libraries" (Spiro, 2009). Van der Vaart (2010: 5), in turn, sees the concept 'collaboratory' as synonymous with a VRE and defines it as a "web-based collaboration environment for researchers." She does, however, make a distinction between the phrase ‘collaboratories’ or ‘collaboratory projects’ to describe actual collaboration activities between researchers in their subject area, and the phrase ‘collaborative environments’ for the software. The Dutch SURFfoundation defines a collaboratory as “a virtual research environment that enables researchers based in different locations to work together and share their knowledge and facilities, thus enriching and speeding up both national and international research” (SURFFoundation, n.d.). Bos et al. (2007: 656), in turn, define a collaboratory as “an organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artefacts, and tools required to accomplish research tasks.” A wider-ranging definition though is provided by Cogburn (2003: 85-86). He defines a collaboratory as “more than an elaborate collection of information and communications technologies.” He sees it as “a new networked organizational form that also includes social processes, collaboration techniques, formal and informal communication, and agreement on norms, principles, values, and rules.”

Most collaboratories initially focused on the natural sciences, for example Nano Hub, Space Physics and Astronomy Research Collaboratory (SPARC) and Biomedical Informatics Research Network (BIRN) (Spiro, 2009). The scope has, nonetheless, been broadened to include the humanities (Spiro, 2009). In the humanities, the concept can be described as “a network of individuals and institutions inspired by the possibilities that new technologies offer us, a national coordinating body and knowledge resource for digital humanities scholarship, “an interdisciplinary research unit, a collaboration among supercomputing centres to support humanities scholars in their use of HPC, a university-based team that supports teaching and research by bringing together computing and the humanities, as well as a scholarly web space that supports collaborative annotation and publication” (HASTAC, n.d.; Spiro 2009). In other words, the concept ‘collaboratory’ acquired additional meanings, which refers to “a new networked organizational form that also includes social processes; collaboration techniques; formal and informal communication; and agreement on norms, principles, values, and rules” (Cogburn, 2003: 86).

The ‘collaboratory’ concept nonetheless appears to be supplanted by the concept VRE, “to refer to online collaborative spaces that provide access to tools and content” (Spiro, 2009).

Another related term to VREs is WRSS.

2.2.7 Web-Based Research Support Systems (WRSS)

Research Support Systems that are done via the web are called WRSS. WRSS according to Tang et al. (2003: 21) is a special type of Intelligent Web Information System (IWIS) that “can be viewed as a concrete research area of Web Intelligence (WI). WI in this context is concerned with the exploration of the “fundamental roles as well as practical impacts of Artificial Intelligence (AI) and Advanced Information Technology on the next generation of Web-empowered products, systems and services” (Zhong, Liu and Yao, 2003: 1). WRSS aims to develop “new and effective tools for research institutions, researchers and scientists” so as to support their research activities and assist them in the improvement of their research quality and productivity (Tang et al., 2003: 21). Research support systems in general, according to Yao (2003: 601), are designed to “support scientists in finding relevant information, choosing the right tools and producing the effective presentation of research results.”

A WRSS renders support at two levels:

- At the institutional level, the support is closely related to Decision Support Systems, focusing on research management and administration; and
- At the individual level, the system assists researchers during every stage of the research process. The support at individual level is concerned with the integration of existing software systems and tools (Tang et al., 2003: 21-23; Yao, 2003: 601-604).

The combination of these two levels results in a complete model of a research support system.

2.2.7.1 Institutional Level: Research Support For Management Personnel

According to Tang et al. (2003: 23), management services can be concentrated into four areas, and Research Support Services play a role in each of these. These areas are laid out in Table 2.2, which was adapted from Tang et al. (2003: 24). The adaptations included changes in the wording of some of the headings under Areas as well as some changes to the wording of some of the responsibilities and provisions. This was done to make the table more readable and understandable.

Table 2.2: Research Management Responsibilities And Provisions (adapted from Tang et al., 2003: 24).

Areas	Responsibilities	Provision
Comprehensive administration	• Collect and distribute research related information; • Assist researchers in developing and sustaining partnerships with industry, higher education institutions, government, and public and private enterprises; • Maintaining archives; • Fulfil research plan, policies and strategy; • Provision of research training; • Scrutinize and advocate ethical practices in research.	• Policies and strategies structure; • Planning report; • Training opportunities; • Research ethics; • Scientific and technological archives.
Management of Projects	• Provision of information on funding opportunities; • Coordinate proposal and application for projects; • Negotiate contracts; • Oversee the progress of projects; • Systematise result evaluation.	• Knowledge about funding opportunities; • Advice on proposal; • Yearly management report.
Management of results	• Provision of research and development statistics; • Intellectual property protection and management; • Technology transfer; • Publication/dissemination of results; • Arrange sponsors for seminars.	• Marketing intelligence; • Identification and exploitation of intellectual property; • Annual research and development statistics report; • Publication.
Management of Finances	• Management of research funding and grants; • Producing certified financial compliance for individual grants and contracts; • Expense control.	• Project expense report; • Financial final accounts report.

A research support system as stated in Tang et al. (2003: 23) should furnish "managers and researchers with services, including information services, sharing resources, and collaborative work support," and it should also be easily accessible. A WRSS should furthermore realize the requirements of a research office, for example:

- Helping researchers to effectively and efficiently identify funding opportunities, and prepare grants proposals and contracts;
- Providing researchers with information retrieval support that will assist them in finding their interested information efficiently;
- Providing public resources sharing, such as data, computing capacity, programming and testing environment, experiment condition, etc.;
- Assisting research managers to effectively handle administrative affairs (Tang et al., 2003: 23-24).

2.2.7.2 Research Support For Individual Researchers

Tang et al. (2003: 24-26) use Yao's (2003) framework and Graziono and Raulin's (2000) model to construct a model of research procedures for individual researchers. They then identify seven stages in their model: idea generation, problem definition, procedure-design/planning, observation/experimentation, data analysis, results interpretation, and communication stages. To assist researchers in each of these stages, Tang et al. (2003: 25-26) list the following specific supporting functionalities that have to be considered:

- **Exploring support**: In the early stage of research, exploration has an important role to play. Ways of exploration can include browsing databases, libraries and the Web. The Web makes it possible to track the browsing history. Data collected from this can then be analysed by means of machine learning and data mining tools, providing researchers with useful information and tools (Tang et al., 2003: 25).

- **Retrieval support**: Retrieval support assists with the retrieval of related activities, for example browsing, searching, organisation, and utilisation of information. This is especially of value at the stage where the researcher forms solid ideas and does a literature search to find relevant information (Tang et al., 2003: 26).

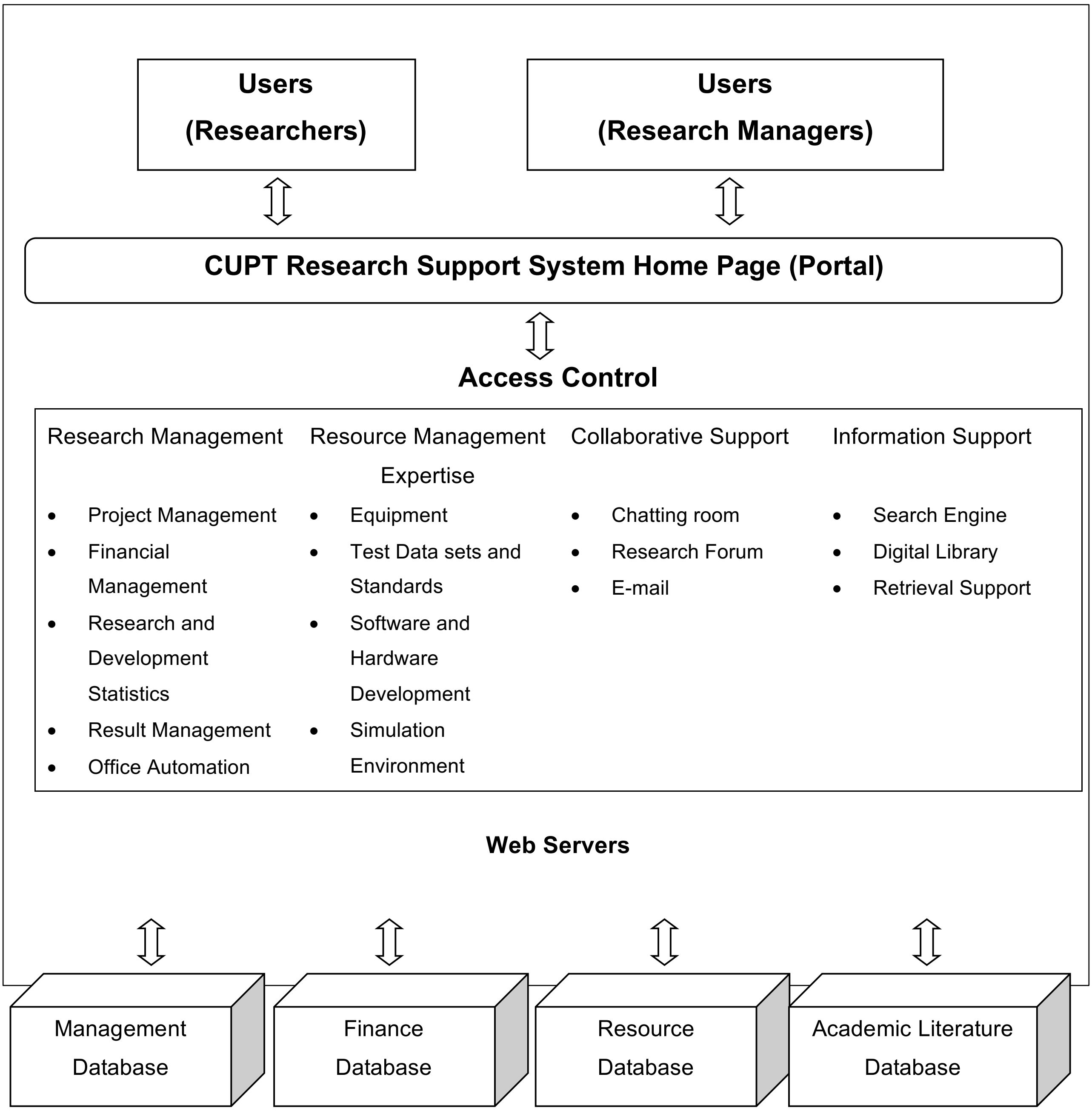
- **Reading support**: In the preparation stage, reading critically and extensively is of the essence, making reading support a necessity. Various software packages make it possible to add bookmarks, to link different sections of an article and make logical connections between different articles. Reading support systems help researchers in finding relevant sources, and also assist them in constructing cognitive maps of the literature read. When combined with exploring and retrieval support systems, machine learning and text mining methods can be used to enable learning from the reading history. Agent technology makes it possible to look for information, and to periodically inform researchers about new information (Tang, et al., 2003: 26).

- **Analysing support**: Helping a researcher find the right tool for a particular problem in analysing data, and assisting him/her in using it, is the role of successful analysing support. Examples of useful tools for analysing support can be computer graphics and visualisation (Tang et al., 2003: 26).

- **Writing support**: Tang et al. (2003: 26) note that there are many writing support software tools similar to word-processing and typesetting software, as well as packages that have additional functions such as spelling-checking, grammar checking and a variety of other agents. Such a writing support system should also include some of the functions named in the retrieval support systems, e.g. a writing support system that can detect relevant articles based on the text compiled by a researcher, and then suggest possible references (Tang et al., 2003: 26).

Tang et al. (2003: 27) discuss an example of a prototype WRSS system (see Figure 2.1), called CUPTRSS (Chongqing University of Posts and Telecommunications Research Support System) - a Chinese system. The aim of the system was to provide researchers with accurate and timely information, to improve research management and to integrate public research sources at the university. The system was mainly designed for management support, and as a test bed or platform for further research in WI technologies. The structure of the CUPT system consists of a multi-layered architecture. The top layer is made up of the different users of the system. The next layer consists of the home page - a portal representing the layers of presentation and business logic. Below these two layers is an access control layer with an authentication module to manage access control. Users access the system through the portal/home page, giving them access to the different databases and functionalities.

Figure 2.1: CUPT Research Support System (Tang et al., 2003: 27)



The CUPTRSS is mentioned by Yang and Allan (2006b) as well as Yang and Allan (2010: 67-68) as an example of a prototype WRSS system. Yang and Allan (2010: 67) emphasize that while there was no implementation of the system, as mentioned by Yao (2004), it provided valuable guidelines (Yang and Allan, 2010: 67). No further changes to this model were found in the literature.

2.2.8 Virtual Research Environments (VREs)

2.2.8.1 What Is A VRE?

A survey of literature revealed a variety of definitions of the concept VREs.

Fraser (2005) views a VRE as “a framework into which tools, services and resources can be plugged.” According to him, VREs “comprise digital infrastructure and services which enable research to take place.” Thanos (2013) agrees with Fraser’s (2005) definition, but expands it by defining a VRE as a “virtual working environment, created on demand, in which communities of research can effectively and efficiently conduct their research activities.”

The UK Joint Information Systems Committee (JISC; 2006) defines a VRE as comprising “a set of online tools and other network resources and technologies interoperating with each other to support or enhance the processes of a wide range of research practitioners within and across disciplinary and institutional boundaries.” According to the JISC (2012), a VRE system can be described as a common flexible framework of resources to “support the underlying processes of research both on large and small scales, particularly for those disciplines that are not well-catered for by current infrastructure.” People then develop and populate this framework with applications, services and resources applicable to their needs. They further describe a VRE as “shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers, [and also] to make use of resources and technical infrastructures available both locally and nationally” (JISC, 2012). Carusi and Reimer (2010: 13) put the JISC definition in other words. According to them, a VRE facilitates collaboration between researchers and provides access to data, tools and services through a technological framework that accesses a wider research infrastructure.

Van Deventer et al. (2009) describe VREs as an intricate part of eResearch and comprising “digital infrastructure and services which enable research to take place within the virtual, multi-disciplinary and multi-organizational partnership context.” They see a VRE not as a product, but as “a framework of integrated and interoperating resources and tools" supporting and enhancing underlying processes of research. They define a VRE as "a mechanism for the creation of a flexible layered architecture of distributed and interoperable resources and tools, [enhancing] the practices and efficiency of individual researchers."

In Dunn's (2009: 205-206) discussion of a VRE, he notes as problematic, the assumption which is often made that 'environment' in this context is the digital equivalent of a research setting that does not use computational networks. This is especially true if one then deduces that it must "refer in some sense to a tangible infrastructure in which research is conducted electronically" (Dunn, 2009: 206). The concept 'e-infrastructure' according to him, further conveys specific assumptions about the capacity it gives researchers, and the scholarly work it enables. The 'research' concept in a VRE, however, is what makes the difference. It links it to a specific class of usage of e-Infrastructure, as well as similar intangibles as mentioned in Unsworth's (2006: 6) definition of cyberinfrastructure, namely an "intangible layer of expertise and the best practices, standards, tools, collections and collaborative environments that can be broadly shared across communities of enquiry."

Another definition of a VRE comes from the DFG, who defines it as "a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information" (translation of the DFG definition of 'Virtuelle Forschungebung' by Carusi and Reimer, 2010: 14). Keraminiyage, Amaratunga and Haigh (2009a: 59), in turn, define a VRE in its simplest form as "a set of web applications intended to enable collaborative research activities beyond geographical barriers," while Leonardo, Castelli and Pagano (2009: 239) define VREs as "providing collaborative frameworks enabling scientists to produce and exchange results with peers around the globe and in cost-efficient manner."

The idea of supporting collaboration is expanded upon by Candela (n.d.: 1), who relates the concept of a VRE to the concept of a Community of Practice (CoP). He regards a VRE as an overarching concept with the following distinguishing hallmarks:

- It is a web-based working environment;
- It is customised to support the needs of a CoP (Lave and Wenger, 1991: 29);
- It is anticipated that it will provide a CoP with the whole spectrum of commodities needed to achieve the community's goal(s);
- "It is open and flexible" with regards to the total service offering and lifetime"; and
- "It promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution" (Candela, n.d.: 1).

Voss and Procter (2009: 175-176) point out that the concept of VREs is still evolving and define VREs as being synonymous with other concepts such as collaboratories, cyberenvironments and science gateways. Wusteman (2009: 170) points out that there is a misuse of terms creeping in, for example the increasing tendency to describe digital libraries as VREs or collaboratories, even though a VRE is more than a digital library, or even a portal to a range of digital activities.

By extracting some of the core elements from above-mentioned definitions, the following definition of a VRE can be compiled:

> A VRE consists of a common, flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts etc.) interoperating with each other, can be plugged, to enable collaboration and to support and enhance large and small scale processes of research, which are often performed by researchers in multidisciplinary contexts, within or across organisational and geographical boundaries.

It is also important to note that the concepts 'platform' or 'framework' is used interchangeably by different authors to describe a VRE (e.g. Fraser, 2005; JISc, 2012; DFG as quoted by Carusi and Reimer, 2010: 14; Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London). For this study, the concept 'platform' includes any hardware or software upon which software applications or services can be built and run (Bigelow and Rouse, 2016; Jamison, Bortlik and Hanley, 2013; Martin, 2014). Separately, "a software framework is used to assist in facilitating software development by providing generic capabilities that can be changed or configured to create a specific software application" (Jamison, Bortlik and Hanley, 2013).

The aim of using a VRE is highlighted next.

2.2.8.2 The Aim Of A VRE

Literature reveals different viewpoints on VRE aims. Voss and Procter (2009: 176) regard the aim of a VRE as providing "an integrated environment that supports a community of collaborating researchers." In other words, a VRE brings together previously separate tools that are needed to do research and to collaborate - aspects integral to a researcher's work. JISC's website on its VRE programme (Phase 1) describes the aim of a VRE as to render assistance to researchers in all disciplines in managing the increasingly complex series of tasks involved in carrying out research. According to JISC, the aim of VREs is "to support e-researchers in their day to day work by providing collaboration functions alongside other tools", for example portals, hardware and scientific equipment, repositories, knowledge management tools, library resources and common desktop applications (Virtual Research Environments: what is a VRE? 2011). Van Deventer et al. (2009) list some of the processes VREs aim to support: funder identification, proposal writing, research administration, communication between all participants, desktop research, data production, data retrieval, data analysis, visualisation, collaborative production of research outputs and project management.

Definitions and aims of VREs provide an overview of what VREs are, but for greater clarity, the different characteristics of VREs will now be discussed.

2.2.8.3 Characteristics Of A VRE

A survey of literature (Allan, 2009: 111; Carusi and Reimer, 2010: 19, 23; Dunn, 2009: 206, 208; Fraser, 2005; Keraminiyage, Amaratunga and Haigh, 2009: 62; Ochem, 2008: 13; Voss and Procter, 2009: 176; Van Deventer et al., 2009; Wilson, et al., 2007: 290; Yang and Allan, 2006a: 453-454) has shown that VREs can have the following characteristics:

- VREs are typically project driven (Dunn, 2009: 206);
- VREs are designed strategically rather than responsively or incrementally (Dunn, 2009: 206);
- A “key characteristic of a VRE is that it facilitates collaboration amongst researchers and research teams providing them with more effective means of collaboratively collecting, manipulating and managing data, as well as collaborative knowledge creation” (Brown, 2012), which brings it in line with the topic that this study investigates;
- A VRE normally has a web-based front end (or portal) that enables clients (researchers) to access the VRE via a web browser using a personal computer (PC) or mobile devices such as cell phones and tablets (Yang and Allan, 2006a: 454);
- A VRE can be described as a one-stop shop where researchers can obtain data and global information pertinent to their research with suitable “semantic support and contextual services for discovery, location, and digital rights management” (Yang and Allan, 2010: 68);
- A VRE has “more in common with a ‘Managed Learning Environment’ (the sum of services and systems which together support the teaching and learning processes in an institution)” than with a VLE (Fraser, 2005);
- A VRE can be constructed on top of existing applications such as VLEs (also called ‘e-Learning systems’) (Keraminiyage, Amaratunga and Haigh, 2009: 62; Voss and Procter, 2009: 176; Yang and Allan 2006a: 453);
- VREs are the products of “joining together new and existing components in support of as much of the research process” as possible for any activity (Fraser, 2005; Wilson, et al., 2007: 290);
- VREs can be used for analysis and processing of data, annotating data collaboratively, and sharing of data with peers (Carusi and Reimer, 2010: 19);
- VREs also enable inter-disciplinarity, by bringing data and approaches from different disciplines together to “create new research findings” (Carusi and Reimer, 2010: 23, Fraser 2005);
- A VRE can provide researchers with new forms of data and challenges to analysis (Wilson et al., 2007: 290);

- Tools/components created for one subject-based presentation of a VRE can potentially be made available to and plugged into other subject-based presentations (Fraser, 2005);
- A VRE can consist of a group of web applications (Keraminiyage, Amaratunga and Haigh, 2009: 62);
- A VRE can be technology-driven, but preferably demand-driven, which will ensure that they are end-user focused (Keraminiyage, Amaratunga and Haigh, 2009: 62);
- “A VRE system should be able to act as communication platform” (Yang and Allan, 2006a: 453; Wilson, et al., 2007: 290);
- A VRE system should be able to support administrative tasks involved in project management, for example “risk assessment, progress monitoring, financial monitoring and task assignments” (Yang and Allan, 2006a: 453);
- A VRE can provide the means for engagement between researchers, policy-makers and practitioners (Wilson et al., 2007: 290);
- VRE systems should be as flexible as possible because user requirements are constantly changing (Yang and Allan, 2006a: 454);
- VREs can follow a three-tier or multi-tier (n-tier) architecture, where web portals can act as the presentation layer, with business logic and data layers behind it (Yang and Allan, 2006a: 454; Allan, 2009: 111). No formal definition of business logic is given in the literature, but it is generally seen as the mid-layer of a web application, with its main components being business rules and workflows (Ochem, 2008: 13; Business logic, 2015). A business rule is seen “as a specific procedure,” while a workflow contains the tasks, the procedural stages, the necessary “input and output information, and tools” required for each stage. Business logic outlines the sequence of actions related to “data in a database to carry out the business rule” (Business logic, 2015);
- A VRE should have the following three components: a recording process (capturing data), clear ownership (through authentication) of the data, and a focus on a specific question or topic – to be formally expressed and documented, “in order to meet the standards of peer-review, research quality assessment, and funding success, that non-digital research is subject to” (Dunn, 2009: 208);

- “A VRE should provide an effective personalised access point to information, experts, knowledge, collaboration tools and computational resources” (Van Deventer et al., 2009).

Yang and Allan (2010: 68) provide a further list of components that researchers would expect in a VRE:

- “Mechanisms for discovering scientific data and linking between data, publications, and citations”;
- VRE discovery services that operate parallel and next to “broad-search services such as Google”;
- Collaboration technologies that can “facilitate joint uses of its services”;
- Web 2.0 and semantic web technologies;
- Facilities for publishing;
- Facilities that can make content available from personal and group information systems;
- Embracing mail list servers and archives;
- Protocols and standards to facilitate exposing its services to an array of “user interfaces, including portals”;
- Provide enhanced access to commercial sources and be interoperable with proprietary software;
- Provide highly customizable interfaces;
- Provide training for users, and
- Awareness of what is available.

The concept of VREs has now been discussed to give greater clarity, but how does the concept differ or compare with other similar terms found in literature?

2.3 DIFFERENCES AND SIMILARITIES WITH THE CONCEPTS CYBER-INFRASTRUCTURE, CYBERENVIRONMENTS, COLLABORATORIES AND SCIENCE GATEWAYS

Fraser (2005) regards the concepts of VREs and cyberinfrastructure/e-infrastructure for the most part as synonymous, but differentiates between the terms as follows. A VRE, according to him, “presents a holistic view of the context in which research is taking place," while cyber-/e-infrastructure "focuses on the core, shared services over which the VRE is expected to operate." In other words, VREs include cyber- and e-infrastructure. Another aspect Fraser (2005) emphasizes is the importance of integrating the VRE "with existing research infrastructure and services." A VRE will in other words include current research infrastructure and services as well as new infrastructure. Furthermore, VREs, according to Fraser (2005), "arises from and remains intrinsically linked with the development of e-science." He notes that the concept of a VRE further contributes to broadening the prevailing e-Science definition, where e-Science is described as "grid-based distributed computing for scientists with huge amounts of data," to a definition that includes "the development of online tools, content and middleware within a coherent framework for all disciplines and all types of research" (Fraser, 2005). Voss and Procter (2009: 175), similar to Fraser (2005), also consider the term VREs as synonymous with cyberinfrastructure, but in addition regard VREs as synonymous with 'collaboratories' and 'science gateways'. Carusi and Reimer (2010: 14), in turn describe cyberinfrastructure as referring to all the aspects of the digital side of research infrastructure, and VREs as the interface to that infrastructure. This corresponds to Basney et al.'s (2011) description of science gateways as convenient interfaces to cyberinfrastructure. A VRE and a science gateway can thus be seen as synonymous.

The concept of WRSS is synonymous with 'web-based VREs' in that "they both improve research support and quality" by contributing RSS and providing collaborative work support (Yang and Allan, 2006b; Yang and Allan, 2010: 68). A 'web-based VRE' could therefore be described as a type of WRSS.

The Dutch SURFfoundation regards the concept 'collaboratory' as totally synonymous with a VRE, and even describes a collaboratory as a VRE in their definition of the concept (see 2.2.6) (SURFfoundation, n.d.). Spiro (2009) indicates that the collaboratory concept appears to be supplanted by the concept of VREs. However, the concept used seems not to be important, according to Carusi and Reimer (2010: 15), and "the understandings associated with the terms VRE, collaboratory and gateway are converging on a set of characteristic features," which can include: access to data, tools and resources; co-operation or collaboration with other researchers at the same or other organisations; co-operation at inter- and intra-institutional levels; or "preserving or taking care of data and other outputs."

For the purpose of this study, the author used the concept of e-Research as a framework from which to investigate the research topic. E-Research rather than e-Science was preferred, as it was seen as a form of scholarship that is broader, covering all ICTs that support researchers in their research process, including all forms of non-computational e-Science. Furthermore, as indicated in 2.2.5.1, there are various approaches to e-Research; however, this study followed the social sciences approach to e-Research, including CMs, IS, SOA and whole process approaches. These sub-approaches each have attributes that were deemed valuable for this study.

The discussion on the computerisation movement approach in 2.2.5.1 showed that it is a "kind of movement whose advocates focus on computer-based systems as instruments to bring about a new social order" (Kling and Iacono, 1998: 228). The discussion also mentioned the advance towards the development of new information infrastructures for research, and the application of these in varied ways across scientific institutions, research fields and disciplines. To understand and explore the application of these technologies, especially in a VRE, and their effect in societal or organisational transformation, the CM approach can be valuable as a framework. The discussion in 2.2.5.1 on the IS approach revealed the importance of it for the interaction between ITCs and human activities systems (e.g. research). It was shown that this approach clarifies and formalises domains of human activity, and creates interventions by IT-based systems in those domains. This relates to the VRE concept where IT-based systems, in this case "a set of online tools and other network resources and technologies interoperating with each other," have an impact on research practitioners conducting research (human activity) (JISC, 2006). As indicated in 2.2.5.1, a SOA approach to VREs will ensure that they are flexible enough for dynamic user needs. The whole process approach as discussed in 2.2.5.1 also looks at issues such as information processing, information exchange, information utilisation and information management in e-Research, which includes the development of demonstrator models "to illustrate how the process will work in practice" (Paterson et al., 2007: 128) - something that would be valuable later in the empirical part of this study.

The researcher decided on the usage of the VRE concept, rather than science gateways or collaboratories, which are synonymous to VREs. The VRE concept is used in various regions across the world (e.g. UK, Europe, Australasia, South Africa; see Chapter 3 for an in-depth discussion), whereas the science gateway concept was seen to be used extensively in the USA and a few countries outside the USA. The collaboratory concept was also shown to be supplanted by the VRE concept, in the literature of the Dutch SURFfoundation and a piece by Spiro (2009). Finally, the concept cyberinfrastructure was used to delineate the core, shared cyber- and e-services over which the VRE is expected to function.

2.4 SUMMARY

In this chapter, a conceptual overview was given as background in order to position the concept VRE. This was done through a description of the concepts of e-Science, cyberinfrastructure, science gateways, and cyberscience. This was the followed by a discussion on the concept of e-Research (VREs being an application of the e-Research field) and the various approaches to e-Research. Other related concepts such as collaboratories and WRSS, as found in the literature, were also discussed. The concept of VREs was then discussed - definitions, aims and characteristics. Differences and similarities of the concepts cyberinfrastructure, cyberenvironments, collaboratories and science gateways to VREs were also highlighted, followed by a discussion on the concepts used for the purposes of this study.

In the next chapter, the current state of VREs as well as components of and tools used in VREs are discussed.

CHAPTER 3
VIRTUAL RESEARCH ENVIRONMENTS AND THEIR COMPONENTS AND TOOLS

3.1 INTRODUCTION

In an increasing technologically developing environment, VREs have become an attractive choice in solving progressively complex research challenges, and researchers in various countries across the world have opted to create or use VREs. Researchers working on a research problem go through a research process to come to a final conclusion. The research process can also be represented as a research cycle, consisting of various iterative stages. By using a VRE, researchers should be able to bolster each of these stages.

This chapter is initiated by giving an overview of the developments of VREs across the world by concentrating on their development in four countries, followed by a discussion of the differences and similarities in the programmes of these countries. The concept of research cycles as well as the various VRE components as found in the literature, is discussed next. Thereafter, the research components as well as possible VRE tools are matched to the research cycle stages. Insights gained from VRE developments internationally, the research cycle concept, the VRE components, as well as VRE tools are then brought together in a possible VRE conceptual framework.

3.2 AN OVERVIEW OF THE DEVELOPMENT OF VREs ACROSS THE WORLD, FOCUSING ON THE UNITED KINGDOM, THE USA, THE NETHERLANDS AND GERMANY

Many countries across the globe are engaged in developing VREs, for example the UK, the USA, the Netherlands, Germany, Australia, Japan, India, Brazil and South Africa (Carusi and Reimer, 2010: 12). Terms used in the different countries may vary though: sometimes the concept VRE is used; sometimes the concept collaboratory; sometimes the concept Science Gateway; and sometimes the concept Virtual Laboratory (Carusi and Reimer, 2010: 12; Wilkins-Diehr, Barker and Gesing, 2016). In recent years up to the conclusion of this study, there have also been regional and/or international VRE initiatives. These are: the e-Infrastructures for VREs under the EU Horizon 2020 programme (e.g. VI-SEEM, MuG, OpenDreamKit, BlueBRIDGE, VRE4EIC, West-Life and Sci-GaIA); the Virtual Laboratories programme under Nectar in Australia; the CANARIE programme focusing on Research Platforms and research software services in Canada; the Science Gateways Community Institute in the USA; the International Coalition on Science Gateways (ICSG); and the Research Data Alliance VRE Interest Group (VRE-IG) (Wilkins-Diehr, Barker and Gesing, 2016).

When investigating the VRE landscape around the world, Carusi and Reimer (2010: 16) found that strategies regarding VREs or similar programmes fall into three main categories:

- Category 1 – dedicated VRE or similar programmes;
- Category 2 – programmes that do not see themselves as specifically advancing VREs, but where there is an overlap with definite VREs or VRE-like programmes; and
- Category 3 – programmes that do not aim for anything like VREs.

The researcher selected four countries as major examples of Category 1, namely the UK, the USA, the Netherlands, and Germany. These countries were selected because they are representative of different VRE approaches or models used across the globe. Programmes in these countries, though each unique in their own way, share a relatively similar vision of key elements of VREs, as shown above, and they are specifically aimed at facilitating the shared use of digital infrastructure by researchers through the provision of shared environments. An overview of these programmes is given below.

3.2.1 United Kingdom

The UK's VRE programme is funded and driven by the JISC a sub-committee of the HEFC (Higher Education Funding Committee for England). JISC's Innovation Group has an e-Research team specifically tasked to look at the following: grid computing, data management (data inside the research process), access management, identity management, etc., as well as collaborative technologies such as VREs (Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London).

Below is a summary of the UK VRE Programme from 2004-2011, as provided in a PowerPoint presentation by Frederique van Till and a PowerPoint presentation by Christopher Brown, both stationed at JISC (Brown, 2012; Van Till, 2005).

Table 3.1: Summary Of The VRE Programme Approaches In The UK: 2004-2011

Phase	VRE1	VRE2	VRE3
Period	2004-2007	2007-2009	2009-2011
Number of projects	15 projects	4 pilots	10 projects
Focus	Technology focused	User- and research practice focused	Broadening use, cross- institutional & discipline
Type	Experimental	Developmental	Developmental
Approach	Diverse design and development approaches	Unified design and development approaches	Diverse design, challenge- and community- driven
Solution	Stand-alone solutions	Integrated pilots	Focused on tools, frameworks and interoperability

JISC's VRE Programme originally started in 2004 with **Phase 1**, which included fifteen projects divided into different strands (JISC, 2014f). All these projects looked at various facets of VREs. The strands and projects according to Brown (2012) were:

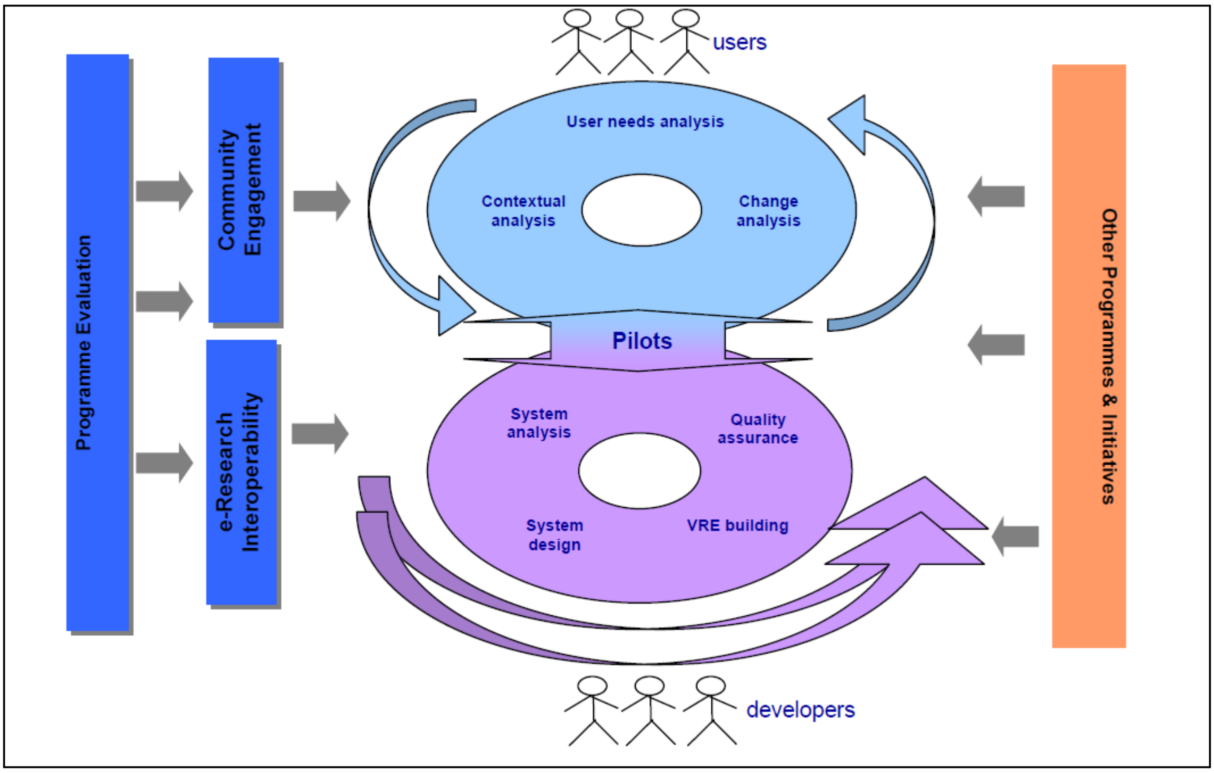
- **Strand 1** – This strand included larger scale projects, and the deployment of VRE demonstrators based on existing frameworks, such as Sakai or Open Grid Computing Environment (OGCE).
- **Strand 2** – This strand included projects that were tasked to identify and add functionality (in the form of tools and services developed in other projects), which at that point had not been integrated into the existing framework architectures.
- **Strand 3** – This strand included projects that looked at the development and deployment of lightweight, proof-of-concept VRE demonstrators appropriate to the needs and skills of specific communities.
- **Cross-Strand** – These were projects that stretched across the different strands.

- **Strand 4 –** This strand consisted of a formative evaluation project assessing the programme according to the following:
  - Establishing how effectively the selected projects were meeting the aims of the programme;
  - Gathering and disseminating best practice;
  - Identifying gaps;
  - Raising awareness of the programme and stimulating discussion on VREs in the community; and
  - Forming an advisory group representative of all sectors of the research community to make recommendations for further work. The formative evaluation was conducted by a team of consultants from the Tavistock Institute in the UK.
- **Strand 5** – VRE Tools and Resources Interoperability Project.

In Phase 1 of the VRE programme, the focus was mainly on experimenting and technology, while the design and development of each of these VREs were very diverse, with stand-alone solutions. The hope was to bring all the facets in these VRE projects together into one VRE solution in a similar manner as VLEs, using shelf-ready tools such as SharePoint, Blackboard, Sakai, Moodle or uPortal; however, the results showed that the fifteen projects had very distinct and different needs with regards to infrastructure and resources. Some used portal technologies, others used Sakai, and a number of others used general institutional web-based tools, which made it difficult to bring them together into one standardised solution (Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London). Frederique van Till from the JISC suggested during an interview in 2010 that it might be possible to create, or use a centralised framework or standardised platform (which can be transferrable), onto which people can build their own tools (Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London).

An evaluation report on Phase 1 was done by the Tavistock Institute in the UK, and results showed that in order to find a technology solution that would work for a project, projects needed to stay close to the users, and start with their questions. This in turn led to the development of the Figure 8 Development Model of Participative Design and Development, which formed the basis for Phase 2 of the VRE Programme.

Figure 3.1: The JISC Figure 8 Model of Participative Design and Development (Van der Vaart, 2010: 27)



Phase 2 of the VRE Programme consisted of four pilot (demonstrator) projects. The focus in this phase was more on the user and research practice utilising unified design and development approaches, and investigated the possibility of bringing everything together into one integrated pilot solution. Developers and users were brought together in a participatory design process, a user needs analysis was done, and this was then contextualised. Something was then built and tested and the process was repeated until they found the best solution. The four projects were: the VERA (Virtual Environments for Research in Archaeology) project; CREW (Collaborative Research Events on the Web - a merger of IUGO and MEMETIC from the 1st round of VRE projects); the SDM (Study of Documents and Manuscripts) VRE; and myExperiment (Interview with F. van Till and M. Dovey, JISC on 1 June 2010 at the HEFC Building, London). Phase 2 was followed by Phase 3 of the VRE Programme, during which JISC funded ten projects divided into three strands/components looking at frameworks, tools, and interoperability, respectively. Four of the projects looked at frameworks (determining how to really build the right framework; deciding what were the most important lessons learned; and investigating how to collaborate); five looked at interoperability; and one looked at tools.

The focus in Phase 3 was to utilise a diverse design, and a challenge- and community-driven developmental approach, to broaden the use of the VREs across institutional and disciplinary borders. The projects in the three strands were:

- Strand 1 – VRE Tools
- Strand 2 – VRE Frameworks
- Strand 3 – VRE Interoperability

Common to all three of the phases of JISC's VRE Programme was a focus on collaboration, support for small- and large-scale research, as well as support for single- and multi-disciplinary research (Brown, 2012). In September 2011, the JISC VRE programme became part of the Digital Infrastructure Research Programme, which focused on assisting researchers and research groups to collaborate and build communities and to exploit e-infrastructure that would give them access to computational resources, storage and platforms - tools that they could use to share and analyse data quickly (JISC, 2014b; JISC, 2014c). The Digital Infrastructure Research Programme ran from 2011 to 2013, and consisted of two strands: Research Tools and Research Support (JISC, 2014c). The Research Tools strand built on the work undertaken in the VRE and Research Infrastructures programme and funded the following activities: a National Grid Service, the exploitation of infrastructure for research (which included distributed computing capability, cloud, visualisation, data mining, semantic services, linked data and geospatial tools), a Virtual Laboratory for the Future (which included hybrid environments / reality, mobile interfaces and new interaction models), and research collaboration and communications (consisting of bridging institutions, research groups, citizen science, scholarly communications, dissemination of research / research impact within the research community, and public outreach) (JISC, 2014d). The Research Support strand focused on supporting UK researchers to take advantage of the opportunities afforded to them through information technologies, and included the following activities: researcher training, institutional ICT support for research by providing models and guidance, research and developer triage, and the JISC advance for research and VRE materials (JISC, 2014e). The Digital Infrastructure Research Programme ended in 2013, which effectively meant that funding provided by JISC for further research on VREs in the UK ended.

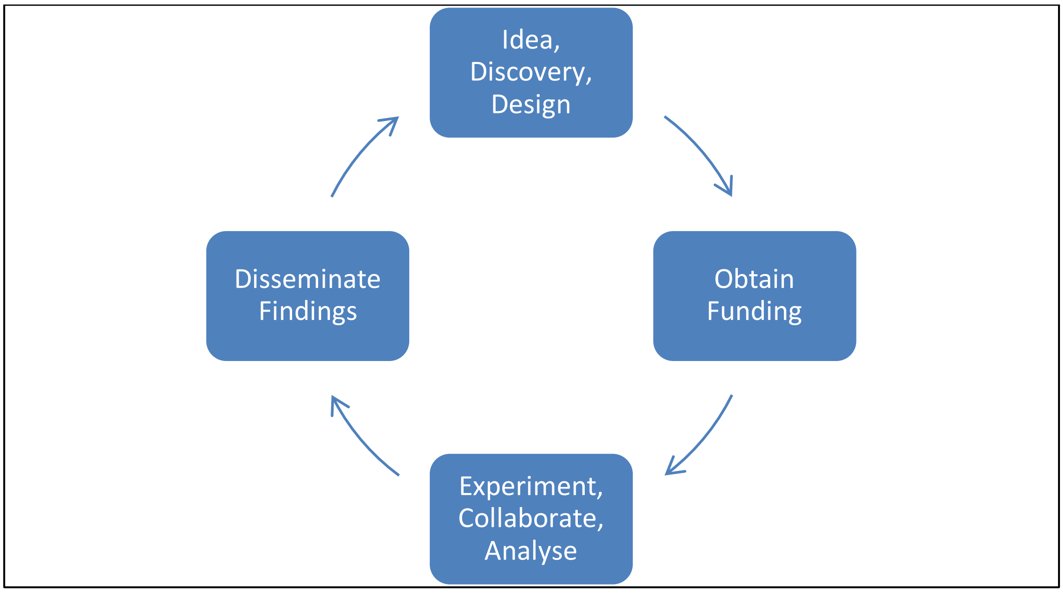
In 2016, the Research Data Alliance (RDA) created a VRE Interest Group to build on the valuable work that was done through the JISC VRE programme, Science Gateways programme in the USA, as well as Digital Laboratories programme in Australia (Glaves, 2016: 3). At the end of 2016, JISC launched a co-design challenge to the research community, to investigate the needs requirements for a next generation research environment (Brown, 2017). This work was seen as the discovery phase of the Next Generation Research Environments (NGREs) project (Brown, 2017). The aim of this project was to "define the future of research environments and to determine how such environments can support the current and future needs of researchers" (Hammond, 2017: 3). A report on this project was presented in early 2017 and revealed that researchers "saw NGREs as either being more capable" VREs (in other words focusing mainly on the "execution of research" and the "collection and sharing of data"), or comprising a much broader scope that covers the "entire research lifecycle" (Hammond, 2017: 6). This broader scope would, for example, be equivalent to combining the functionalities of a complete VRE with a complete CRIS (Current Research Information System), and with an authoring platform (Hammond, 2017: 6). As a result of the report JISC decided to pursue a number of actions:

- Continuously engage international groups working on developing VREs and promote the adoption of concepts, standards and identifiers that are of value to JIS members;
- Ensure that JIS services are developed in a manner that would increase the possibility for integration of these and access to these via APIs and standard interfaces;
- Deliver JISC services to disparate stakeholders in such a manner that it would increase awareness of these in the research community; and
- Investigate the actions that would be needed for a closer integration of active data and archival research data within JISCs Research Data Shared Service, and the actions that would be needed to integrate research data with administrative data, and test this within the JISC test environment, called the 'University of JISC' (Brown, 2017).

Simultaneous to the JISC VRE Programme, the British Library and the Technical Computing Group at Microsoft started a joint venture to develop the Research Information Centre (RIC). The RIC ran on the Microsoft SharePoint Platform and provided a set of core functionalities that covered all facets of the research lifecycle (“a high-level view of the cyclic nature of research”) (Barga, Andrews and Parastatides, 2007: 31; Carusi and Reimer, 2010: 89). The RIC divided the Research Cycle into 4 main components:

- Idea discovery and design;
- Obtaining funding;
- Experimenting, collaborating and analysing; and
- Dissemination of findings.

Figure 3.2: Research Lifecycle (Barga, Andrews and Parastatides, 2007: 31)



The core functionalities of the RIC could be used to build domain specific VREs, into which additional modules could be added (Carusi and Reimer, 2010: 89). The RIC furthermore aimed to lessen the amount of time researchers had to spend on administrative tasks. It also aimed to offer easy access to relevant information and information sources, to facilitate networking, and to preserve/curate not only the project outcomes, but the whole process of research (research cycle) (Carusi and Reimer, 2010: 89). Key areas that were addressed by the RIC were content and knowledge management, social networking and online collaboration. Users could create templates for projects and set up specific project sites based upon those templates. Features offered included “access control; workflows; sharing and annotation of resources; RSS feed integration; federated search over domain-specific literature sources and a full-text search over local resources; blogging; wikis; networking; creation of project groups; bibliographical support; and archiving of project sites" (Carusi and Reimer, 2010: 90). The RIC could be "deployed as an institutional VRE environment," which "could support the management of projects and facilitate sharing of information across the institution." At the same time, it was able to provide "researchers with a domain and a project specific environment" wherein they could add additional resources (Carusi and Reimer, 2010: 90). The RIC could also be set up in a shared hosting situation, for example as a service offered through the cloud (Carusi and Reimer, 2010: 90).

The RIC VRE project was discontinued in 2013, because many of the features that were implemented during the VRE project had been made obsolete by native features that were eventually released in Microsoft SharePoint Server 2010, together with VRE toolkits for SharePoint, which were developed by the British Library, Oxford University, the University of Southampton, University of Delhi, LaTrobe University, and SoftEdge Systems (Research Information Centre Framework, 2015).

3.2.2 USA

It was very difficult to obtain an overview of developments in the USA because there seems to be no co-ordinated programme and funding in the area of VREs or similar environments. Nonetheless, it was determined that in 2002, the Science of Collaboratories (SoC) project was launched by CREW (Collaboratory for Research on Electronic Work). The basis of its work was to gather information on collaboratory projects in the USA. Some of the projects identified dated back to the 1960s and 70s, even before the concept of collaboratories was formulated (Van der Vaart, 2010: 26). Other advances in the USA with regard to VREs include the development of Sakai (a VLE used for a number of VRE projects) and the Bamboo project, both funded by the Mellon Foundation (Van der Vaart, 2010: 26).

Literature shows further that the VRE programme in the USA has been mainly driven by science gateways. A science gateway, according to XSED (Extreme Science and Engineering Discovery Environment) (2017c), is "a community-developed set of tools, applications, and data that are integrated via a portal or a suite of applications, usually in a graphical user interface, which is further customized to meet the needs of a specific community." These gateways make it possible for entire communities of users sharing a common discipline to use national resources in the USA through a common interface that is configured for optimal use. It also "fosters collaboration and sharing of ideas" between scientists (XSED, 2017c). From 2001-2011, the Science Gateway Project ran as a sub-project of the TeraGrid Project. This project was funded by the National Science Foundation (NSF). The science gateways aimed to facilitate the use of TeraGrid resources by researchers, and ultimately a greater take-up of TeraGrid and High-Performance Computing (Carusi and Reimer, 2010: 51). The TeraGrid consisted of eleven partners: Indiana University Research Technologies Division, Louisiana Optical Network Initiative (LONI), National Center for Atmospheric Research (NCAR), National Center for Supercomputing Applications (NCSA), National Institute for Computational Sciences (NICS), Oak Ridge National Laboratory (ORNL), Pittsburgh Supercomputing Center (PSC), Purdue University, San Diego Supercomputer Centre (SNDC), Texas Advanced Computing Centre (TACC), and the Argonne National Laboratory (Carusi and Reimer 2010: 51).

In the TeraGrid Project, new projects applying for a science gateway had to go through an open peer-reviewed process of evaluation. By 2010, the TeraGrid Project had consisted of 35 gateways that used TeraGrid as well as other resources, including some cloud-based ones. Carusi and Reimer (2010: 51) found that the gateways were mostly in the natural and physical sciences, with only one in the social sciences. They also found that most of the projects were using computing resources rather than applications geared towards data management, although there were the beginnings of making data available, for example, in expensive petascale simulation. The Science Gateways Project was non-prescriptive about which technology or software to use, and used a bottom-up and user-driven approach; in other words, designing the environments according to the needs of their communities. Resources that could be accessed and utilised through these gateways included workflows, data collections, data analysis and movement tools, resource discovery tools, visualisation tools, and job execution services (Carusi and Reimer, 2010: 51). Carusi and Reimer (2010: 51) identified three instances of science gateways as being the most common:

- Web Portal: a web-browser-based application as interface with users before and TeraGrid services behind;
- Desktop application: an application or suite of applications that forms the interface and run directly on users' personal computers (PCs), while accessing TeraGrid services;
- Grid-bridging gateway: community-run grids devoted to communities' own areas of science, but also extending the reach of their community grid so that users can access and use resources of the TeraGrid.

In 2011, the TeraGrid Project was replaced and expanded by the Extreme Science and Engineering Discovery Environment, a project funded by the NSF (XSEDE, 2012: 23; XSEDE, 2017d). XSEDE currently consists of the following partners: National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign; Center for Advanced Computing (CAC) at Cornell University; Pittsburgh Supercomputing Center (PSC), a joint effort between Carnegie Mellon University and the University of Pittsburgh; San Diego Supercomputer Center (SDSC) at the University of California San Diego; Texas Advanced Computing Center (TACC) at the University of Texas at Austin; Arkansas High Performance Computing Center at the University of Arkansas; Center for Education Integrating Science, Mathematics, and Computing (CEISMC) at the Georgia Institute of Technology; High Performance Computing Center at Oklahoma State University; Information Sciences Institute at the University of Southern California; National Center for Atmospheric Research (NCAR) at the University Corporation for Atmospheric Research (UCAR); National Institute for Computational Sciences (NICS) at the University of Tennessee, Knoxville; Ohio Supercomputer Center; Pervasive Technology Institute (PTI) at Indiana University; Rosen Center for Advanced Computing at Purdue University; Shodor Education Foundation; Southeastern Universities Research Association (SURA); Supercomputing Center for Education and Research at Oklahoma University; Terry College of Business at the University of Georgia; the University of Chicago; and the Argonne National Laboratory (XSEDE: Extreme Science and Engineering Discovery Environment, 2017a). It is led by the University of Illinois's National Center for Supercomputing Applications (NCSA)" (XSEDE, 2017a).

The XSEDE ecosystem include a wide portfolio of resources such as “multi-core and many-core” HPC systems, distributed high-throughput computing (HTS) environments, data analysis and visualisation systems, large-memory systems, data storage, and cloud systems (XSEDE, 2017b). Some of these resources can be accessed through a central XSEDE-managed allocations process, but several resources operated by members of the XSEDE Service Provider Forum are also linked to certain parts of the ecosystem (XSEDE, 2017b).

The Science Gateways project has not changed tremendously under XSEDE. XSEDE (2017c) describes science gateways as “portals to computational and data services and resources across a wide range of science domains for researchers, engineers, educators, and students,” providing, depending on the need, any of the following features:

- “High-performance computation resources;
- Workflow tools;
- General or domain-specific analytic and visualization software;
- Collaborative interfaces;
- Job submission tools; and
- Education modules.”

XSEDE currently lists 33 science gateways and support these in a number of ways, through:

- Provision of support community accounts through XSEDE service providers, which enables gateways to execute scientific applications on XSEDE resources as generic gateways users;
- Allocation of Virtual Machine hosting for science gateways, as well as services relating to them; and
- Rendition of services to assist science gateway providers with the integration of “new and existing science gateways with XSEDE resources” (XSEDE, 2017b).

Other VRE developments in the USA include Alzforum, Schizophrenia Research Forum, and StemBook on the one side, and HUBzero and Open Science Framework (OSF) on the other side. Alzforum, Schizophrenia Research Forum, and StemBook are a cluster of projects in biomedical research, which focus in the first instance on neurodegenerative disorders (Carusi and Reimer, 2010: 61). The funding for these projects come mostly from a philanthropic foundation, and is the result of a collaboration between the foundation, a team of dedicated staff members and consultants, and an interdisciplinary team of bio-informaticians and other biomedical researchers at the Mass General Institute for Neurodegenerative Disease (Carusi and Reimer, 2010: 61). Alzforum is the longest standing of these projects. It was started in 1996 through the provision of funding from a philanthropic organisation that saw the benefits that could be obtained by using the Internet for collaboration in biomedical research, and specifically on Alzheimer's Disease (Carusi and Reimer, 2010: 61).

Alzforum started off as a website that functioned as a type of community newspaper. On the site, paraphrased abstracts of international papers on Alzheimer's disease were published, as well as slides and audio of relevant presentations at scientific conferences. There was also a collection of Milestones Papers in Alzheimer's Disease research and a 'Paper of the Week' feature. The site furthermore facilitated fast informal communications between researchers, for example through live chats and through comments and discussions about papers (Carusi and Reimer, 2010: 61). "Alzforum maintains a range of databases covering biomarkers, brain banks, risk factors, diseases genetic association studies, benign genetic variability, gene mutations, protocols, gene association studies, research models, and therapeutics (Alzforum, 2017). Alzforum also integrates "these diverse sources, by linking primary research articles to related news, papers, databases, and discussions, etc." (Carusi and Reimer, 2010: 61).

Alzforum is described by Carusi and Reimer (2010: 61) as a socio-technical organisation where editors play a key role to promote and moderate conversations and commentary, as well as discovery and integration of information. The system also has a data-driven dynamic system to search and download PubMed citations into a database every night. Semantic tools are being developed to "assist in the identification of hypotheses and related evidence in papers and discussions" (Carusi and Reimer, 2010: 61). A further ongoing development is the "integration of their websites with Web 3.0 functionality," which includes systems that will "enable semantic web applications for representing hypothesis and evidence in scientific discourse" (Carusi and Reimer, 2010: 61). Semantic web applications can also be of great value in analysing content in a VRE.

Another project, similar in functionality to Alzforum, is the Schizophrenia Research Forum (SRF). In 2003, a number of researchers formed a partnership with “NARSAD, the Mental Health Research Association” (now the Brain and Behavior Research Foundation (BBRF)) to form the Schizophrenia Research Forum (SRF) (Schizophrenia Research Forum, 2017a). Through this partnership, funding was obtained from “the National Institute of Mental Health (NIMH) to support the website's development and first years of operation” (Schizophrenia Research Forum, 2017a). Presently “the BBRF provides all the support for the SRF website” (Schizophrenia Research Forum, 2017a). The purpose of the project is to “foster collaboration” between researchers (specifically those working on schizophrenia, those working on related diseases, and basic scientists) through the provision of an “online forum where ideas, research news and data can be presented and discussed” (Schizophrenia Research Forum, 2017b).

The Stembook project site contains a collection of open access chapters dealing with an array of topics related to stem cell research, written by some of the best researchers in the field based at the Harvard Stem Cell Institute and further afield (StemBook, 2013). These chapters were linked to related databases, which made it possible for readers to post comments and discuss entries (Carusi and Reimer, 2010: 61). The project was run in collaboration with a range of other institutions such as the University of Massachusetts Medical School, National Institutes of Health and MassGeneral Institute for Neurodegenerative Disease, as well as entities such as Harvard’s Initiative in Innovative Computing. Its website is still available at https://www.stembook.org/, but activity ceased after 2013 (StemBook, 2013).

HUBzero was initially developed by Purdue University “to support nanoHUB.org”, an online community for the Network of Computational Nanotechnology (NCN). It is now undergirded by a consortium comprising Purdue, Indiana, Clemson and Wisconsin Universities (McLennan and Kline, 2011; HUBzero, 2017a). HUBzero (2017a) describes itself as “an open source software platform” / tool that “hosts analytical tools,” which can be used to “publish data, share resources, collaborate and build communities in a single web-based ecosystem.” In addition, HUBzero contains a powerful content management system, as well as scientific simulation and modelling tools (HUBzero, 2017a). It makes use of hubs and hub-building and researchers and/or institutions have the ability to build and host their own hubs (HUBzero, 2017a). The hubs provide for collaborative development and dissemination of computational models operating in an infrastructure that makes use of cloud computing resources (Mclennan and Kline, 2011). Each hub's tools do not come from the core development team, but from other researchers scattered across the globe, with HUBzero supporting their workflow (McLennan and Kline, 2011). A group within a hub can have a customisable mini website, with features such as calendars, wikis, messaging, blogs, discussion forums and multimedia features such as embedding slides and videos (Purdue University, 2011; HUBzero, 2017a). Authentication and authorisation in groups are possible with different levels of authorisation/access, e.g. 'private', 'available to everyone' and 'available only to group members or invitees' (Purdue University, 2011). HUBzero also has a social networking feature that enables the formation of communities of researchers, educators and practitioners across disciplines, and facilitates communication, collaboration and distribution of research results, as well as education and training (McLennan and Kline, 2011).

HUBzero's functionality has not been mixed together with commercial web software, but has been integrated in a single package. It provides access to, tracking of, and storage of data, and it could also be used to build a web-based repository of models and related documentation, with the added ability to make models operable in a web browser window. It furthermore has built-in features such as a wiki to share ideas and information (McLennan and Kline, 2011). HUBzero operates with open source software such as Debian, LDAP, PHP, Apache HHTP Server, GNU Linux, MySQL, OpenVZ and Zoomla, and its Middleware accommodates live simulation sessions and enables the easy connection of tools to supercomputing clusters and cloud computing infrastructure used to solve large computational problems (McLennan and Kline, 2011; HUBzero, 2017b). HUBzero's simulation tools "are running on cluster or cloud hosts and [are] projected to the user's browser via virtual network computing (VNC)" (McLennan and Kline, 2011; HUBzero, 2017b). Each of these tools operates within "a restricted lightweight virtual environment [by] using OpenVZ," which also manages "access to the file systems, networking and other server processes" (McLennan and Kline, 2011). The system can furthermore route jobs to national resources in the USA, such as the Open Science Grid, the TeraGrid, Purdue University's DiaGrid, as well as other cloud-type systems (McLennan and Kline, 2011). A content management system is used for the publication of tools, while an exclusive HUBzero workspace for developers provides a space where they can create and test their tools in a similar fashion and environment as the published tools, “with access to the visualisation cluster and cloud resources for testing” (McLennan and Kline, 2011; HUBzero, 2017a). This workspace is nothing else but a Linux desktop operating in a secure execution environment, and accessed like any other hub tool, through a web browser.

HUBzero’s Rappture toolkit can be used to turn research modelling and simulation codes into a graphical user interface (web-enabled programmes). Rappture also has a regression tester tool that allows researchers to create test results, which they can run against a large collection of input values, to assist in “verifying that the software is functioning correctly” (Purdue University, 2011; NanoHUB, 2017b). Other functions of HUBzero include a fast searching function, a blog module for personal profiles, blogs for groups, analytics and per-contributor report function, a Twitter feed function, a built-in trouble report, a community forum modelled after Amazon.com’s Askville, and the possibility to use some commercial collaboration tools, e.g. Adobe Presenter, in HUBzero (Purdue University, 2011; McLennan and Kline, 2011; HUBzero, 2017a).

The Open Science Framework (OSF) was developed by the Center for Open Science (COS), a non-profit organisation that was launched through sponsorship from the Laura and Johan Arnold Foundation in 2013, and is still funded through sponsorships, donations and grants (Center for Open Science, 2017). The OSF is a free cloud-based tool that facilitates open centralised workflows through the enablement of different features and outputs of the research lifecycle, such as developing a research idea, designing a study, storing and analysis of collected data, and writing and publishing of papers or reports. The core feature “of the OSF is its ability to develop projects” (Foster and Deardorff, 2017: 203). A project operates as a workspace, which is designed according to users’ preferences and their type of research workflows. The standard project layout comprises a wiki, spaces where files can be uploaded, tags can be added and components (e.g. sub-projects) can be created, and a log of recent activity. The OSF is also collaborative in nature and users can easily add contributors to projects (Foster and Deardorff, 2017: 203). The OSF furthermore has an authentication function so that members of a project can be assigned different level of access, for example read only, read and write, and administrator (Foster and Deardorff, 2017: 203). Users have the ability to assign digital object identifiers (DOIs) and archival resource keys to project, if they are made openly available (Foster and Deardorff, 2017: 203).

The OSF has different licensing options available to users for sharing, such as Creative Commons, MIT, Apache and GNU General Piblict, etc. These licences can be applied to the project as a whole or to specific parts of a project. Another valuable feature of the OSF is that it allows for third-part add-ons or integrations, which can fall into two categories, for example, citation management integrations, for example Mendeley and Zotero, and storage integrations, for example “Amazon S3, Box, Dataverse, Dropbox, Figshare, GitHub, and oneCloud” (Foster and Deardorff, 2017: 203-204). Last but not least, OSF has an ‘OSF for Institutions’ programme that would require additional configuration by the Centre for Open Science and Information Technology staff of the specific institution (Foster and Deardorff, 2017: 204).

Another collaborative development is the creation in 2016 of the Science Gateways Community Institute, a multi-institutional consortium launched with funding obtained from the NSF “to increase the capabilities, numbers and sustainability of science gateways” (Science Gateways Community Institute, n.d.; NSF, 2016). The Institute offers the following services:

- **Incubator**, offering opportunities to “learn best practices” from experts in the field;
- “**Extended developer support**,” offering “direct custom development” assistance;
- “**Scientific Software Collaborative**,” which assists researchers in finding “gateways or software components,” or to promote their own;
- “**Community engagement and exchange**,” rendering opportunities for researchers to “engage with and learn from” the Science Gateways community; and
- **Workforce development**, which assist students or young professionals in developing professional careers in this field (Science Gateways Community Institute, n.d.; Wilkins-Diehr, Barker and Gesing, 2016).

3.2.3 The Netherlands

The VRE Programme in the Netherlands has mainly been driven by SURF, an organisation promoting collaboration among higher education institutions on issues regarding Information Communication Technology (ICT) for education and research

(Carusi and Reimer, 2010: 53). The user-facing division of SURF is called SURFfoundation, and the main technical and development division is called SURFnet. SURFfoundation deals with a wide range of ICT-related areas, including Scholarly Communications. SURFshare was a SURFfoundation project that was pioneered in 2007, and which was committed to the construction of "a common infrastructure that [would] facilitate access to research information" and to the enablement of researchers to share scientific and scholarly information (SURFshare, n.d.). The Dutch higher education institutions, the Royal Netherlands Academy of Arts and Sciences (KNAW), and the Netherlands Organisation for Scientific Research (NWO) were partners in this project until its termination in 2011. It was then taken over by, and is currently run, by SURF's SURFshare Operating Division (SURFshare, n.d.).

The SURFshare project had six major themes, and collaboratories (a similar term to VREs), was one of these (Carusi and Reimer, 2010: 53). The collaboratories theme focused primarily on supporting specific research processes that induced publication, for example, using and re-using research data, the tools to work on those data, archiving, and collaboration; in other words, focusing on the research process rather than the output per sé. SURFnet took responsibility for the development of the technology, but did not prescribe or support a specific technology. A flexible approach was followed to meet researchers' needs. Funded projects were given the freedom to test any environment that would meet their needs, and from this, it emerged that data sharing was an important and central feature of these collaboratories (Carusi and Reimer, 2010: 53). In 2007 and in 2008, calls for tenders for collaboratories were issued by SURFshare. In the 2007 tender projects, the focus was on short-term projects development and implementing environments. In the 2008 tender projects, the focus was on widening and deepening the first experiences of the 2007 projects, for example by increasing experience in more disciplines/institutes, as well as gaining more insight in user experiences, as well as the impact of the collaboratories on their work. Five tender collaboratory projects were funded, namely Collaboratory for Evidence Based Critical Reviews, Hublab, Tales of the Revolt, Testweeklab, and Virtual Knowledge Studio (VKS). Examples of other collaboratory projects not funded by the SURFshare programme are:

- Alfalab, a joint project by five institutes of the Royal Netherlands Academy of Arts and Sciences (DANS, Fryske Akademy, Huygens Institute, Meertens Institute and the Virtual Knowledge Studio), with the aim "to provide an e-Research infrastructure by uniting digital sources and tools for analysis, [and bringing] together Geolab with online tools for georeferencing, annotating and visualising geodata, as well as Textlab with "online tools for co-operative tagging (enhancing) of text data, supported by a supervised learning machine, repositories for data [and also] applications and materials such as index, tutorials, manuals and online dissemination tools" (Van der Vaart, 2010: 23).

- Collaboratory.nl, a research and development project by Novay (Telematica Institute), Philips, DSM, Corus, FEI, and the University of Amsterdam, which ran from 2003-2006, aimed at developing a practical application by integrating technology for remote operation of laboratory instruments with groupware, so as to enable online remote collaboration between researchers /experts and clients in industry. Software for the project consisted of 95% open source software including "portal technology, security software, authentication and authorisation, and collaborative tools" (Van der Vaart, 2010: 24).

- Digital Collaboratory for Cultural Dendrochronology in the Low Countries (DCCD), a project of the Cultural Heritage Agency, DANS (Data Archiving and Network Services) and Utrecht University, which ran from 2008-2010, and which was aimed at "the international standardization of dendrochronological data and metadata, the development of a sustainable and integrated repository of these data, [as well as] unlocking these data for interdisciplinary follow-up research, including the development of a controlled four-language vocabulary based on a number of existing vocabularies" (Van der Vaart, 2010: 24; Digital Collaboratory for Cultural Dendrochronology: an international digital library for dendrochronology, n.d.). A repository was developed in conjunction with DANS for storage, search and retrieval and re-use and is available as freeware (open source software) to members of the collaboratory and the public;

- eLaborate, a project of the Huygens Institute/Royal Netherlands Academy of Arts and Sciences, which was birthed in 2004 with the aim "to realise a collaborative framework for the creation of text editions and textual research in online working environments" (Van der Vaart, 2010: 25). They developed their own framework in which "researchers can work individually or with a group of collaborators on the transcription or edition of a text" (Van der Vaart, 2010: 25). Digital presentations and printing of editions are possible, as well as an adaptable system of categories in order to make it possible to distinguish between the different stages of an edition (Van der Vaart, 2010: 25);

- LabsOnline, a project which was run jointly by the University of Amsterdam, VU University Amsterdam, Fontys University of Applied Sciences, University of Applied Sciences Utrecht and Hague University of Applied Sciences, from 2006-2007. A number of examples of online laboratories in educational settings were developed and the technical and pedagogical implications of these remote laboratories in education were explored. The project had an online registration system for the laboratories and their users, and developed 20 experiments that students could conduct online (Van der Vaart, 2010: 25).

- PARTNER, a project ran by Utrecht University Library from 2006, with the aim of implementing virtual knowledge centres in research groups for an assortment of purposes, e.g. education, research, internships, and involvement of alumni. SharePoint 2007 was used as the basis, with modifications. The University Library provided the programme as well as the project management and implemented the project in close collaboration with the groups involved. The SURFshare EBCR project was one of the projects in this programme (Van der Vaart, 2010: 25).

In 2011, the Netherlands Organisation for Scientific Research (NOW) and SURF formed a unique collaboration to launch the Netherlands eScience Center (NLeSC). This centre was formed in response to a request by the Dutch government "to develop a sustainable coherent and cost-effective eScience environment and e-infrastructure across all scientific disciplines" in the Netherlands (NLeSC, 2015: 9). The need for access to and requirement to manage large volumes of data, the increased complexity of research projects, and the need for effective collaboration between researchers from multiple disciplines and multiple locations, as well as the need for all researchers to become data-scientists, were some of the drivers that necessitated such a Center (NLeSC, 2015: 6). The NLeSC's mission is:

- To "enable scientific breakthroughs" through deploying eScience (which in the context of this study is more aligned to the eResearch concept) "methods, technologies and workflows";
- To facilitate collaboration of researchers from different disciplines in "problem-driven projects";
- To develop "versatile cross-disciplinary eScience tools"; and
- To coordinate "eScience activities" by partnering with partner organisations "nationally and internationally, to identify common challenges," for example in the area of career support for researchers in eScience, and by providing "leadership on issues such as data-stewardship" (curation) and software sustainability (NLeSC, 2015: 11). The Netherlands eScience Center can be accessed at https://www.esciencecenter.nl/.

3.2.4 Germany

The German VRE Programme has mainly been funded and driven by the Deutsche Forschungsgemeinschaft (DFG), the German Research Foundation. The DFG has been interested in virtual research collaboration for a long time and started its first related programme 'Themenorientierte Informationsnetze' (issue-focussed information networks) in 2000. Specific VRE calls, 'Virtuelle Forschungsumgebungen', were issued in 2008 and in 2009 "to support collaborative working across disciplines and over the whole research lifecycle, from collecting and sharing of primary data to analysis, publication and preservation" (Carusi and Reimer, 2010: 46). The DFG welcomed interdisciplinary and international projects, but had one prerequisite, namely that the projects had to be a collaboration between researchers and infrastructure developing institutions (libraries, computer centres, e-Research centres) (Carusi and Reimer, 2010: 47). The calls did not prescribe any particular technology, as different technologies were seen as suitable for different research questions. The DFG nonetheless encouraged the development of new software following open source principles and demonstrating awareness of state of the art and relevant standards. This included support for the development of infrastructure and testbeds. DGF funded two types of projects, namely development projects ('Entwicklungsprojekte'), in other words projects that developed something new, and transfer projects ('Transferprojekte'), which applied existing solutions (Carusi and Reimer, 2010: 46). With regards to technology, however, it is difficult to identify a clear trend, because of the diversity of the projects. The eSciDoc platform tends to stand out, as well as a substantial number of projects that utilised repositories and grid architecture, and a substantial number that used lightweight Web 2.0 technologies (Carusi and Reimer, 2010: 46).

In the 2008 round, 15 bids were received, six of which were subsequently funded in 2009 (Carusi and Reimer, 2010: 46). In 2009, the number of applications for the second round increased significantly to 48, and subsequently, 16 projects were approved in 2010.

The Alliance of Science Organisations in Germany instituted the first phase of the Priority Initiative ‘Digital Information’ in 2008, which stretched till 2012 and expanded on the ideal of having an innovative information environment by focusing on:

- Provision of the widest possible range of “access to digital publications, digital data”, as well as “other source materials”;
- The employment of digital media to develop the ideal environment “for the distribution and reception of publications” that cover German research;
- Ensuring the long-term preservation “of the digital media and contents that have been” collected throughout the world, as well as their “integration in the digital” (virtual) research environment; and
- Supporting “collaborative research” through the utilisation of “innovative information technologies” (Alliance of German Science Organisations, 2008: 1-2).

The priority areas that were identified in this initiative comprised: national licensing, Open Access; a national hosting strategy; research data; VREs; and legal frameworks (Alliance of German Science Organisations, 2008: 2). The Priority Initiative that focused on VREs set a goal of designing a research and development strategy to support researchers as they establish subject-specific and interdisciplinary VREs. The Priority Initiative also set a goal of intensifying and extending some of the projects funded under the DFG programme, called ‘Thematic Networks’ (Alliance of German Science Organisations, 2008: 6). Results from these projects would then, at a later stage, be used in making a decision on how to intensify collaboration within the alliance “regarding the establishment of cross-institutional VREs” (Alliance of German Science Organisations, 2008: 6). During this phase, a working group was set up and a survey was conducted among facilitators of selected VREs to gather more facts about the nature and organisation of VREs (Alliance of German Science Organisations, 2013: 9). The results led to the formulation of a set of guidelines for supporting researchers who want to build a VRE (Alliance of German Science Organisations, 2013: 9).

In 2013, a second phase of the Priority Initiative ‘Digital Information’ was launched, covering the period 2013 to 2017. The second phase has continued work in each of the priority areas that were identified in the first phase, while cross-disciplinary topics are being handled by ad-hoc working groups within a stipulated time limit (Alliance of German Science Organisations, 2013: 9). Four tasks were presented to the working group for VREs:

- **Mapping and analysis**, which comprises the description, standardisation, substantiation and analysing of existing VREs on the basis of results from the survey that was run in the first phase, as well as the Community for Academic Reviewing, Publishing and Editorial Technology (CARPET) project, a DFG project, aiming for the “creation of an infrastructure that allows information exchange and communication by means of VREs”;
- **Transition to permanent operation**, which entails the development of guidelines and recommendations for the “moving of a VRE from the set-up stage to the operational stage” (permanent operation);
- **Legal issues**, which involves the identification of legal issues concerning organisational forms and types (e.g. cross-border usage of nationally funded resources), the development of possible solutions with competent partners, as well as the initiation of their implementation; and
- **Exchange of experiences**, which entails the organising of a range of workshops focusing on practical experience, with the aim to embolden researchers that have been involved in projects, to have conversations about their experiences – especially about critical success factors such as social aspects, acceptance, technology and quality (Alliance of German Science Organisations, 2013: 9-10).

The overview of the VRE projects in the four countries showed some similarities and differences. These will be discussed in the following section.

3.3 SIMILARITIES AND DIFFERENCES BETWEEN THE VRE PROGRAMMES IN THE UK, THE USA, THE NETHERLANDS, AND GERMANY

The similarities and differences in the VRE programmes of the different countries can be listed by using Van der Vaart's (2010: 35-44) topical groupings as a point of departure. These are: organisational, technical, functional, policy/legal/financial, and cultural aspects.

3.3.1 Organisational Aspects

Unlike the other countries' projects, the German DFG required its funded projects to be collaborations between researchers and infrastructure developing institutions, such as libraries, computer centres, and e-Research centres (Carusi and Reimer, 2010: 47). The UK projects showed the importance of including the users of these projects in the design process. JISC made use of a Figure 8 Participative design-process, where users and developers together design a VRE. A user needs analysis, as well as a contextual and change analysis were done among the users, systems were analysed and designed, and VRE pilots were built, keeping quality assurance in mind. This can be seen in the second phase of the British VRE programme, where the focus was on the user and research practice. In the USA, a similar bottom-up and user driven approach was followed with regards to the technology or software used. In the Netherlands, the SURFNet programme followed the same approach. Users were given the freedom to experiment, and also in the German DFG programmes, projects were given the freedom to develop their own technologies, or adapt existing ones.

3.3.2 Technical Aspects

Some of the UK Projects used shelf ready tools like Sakai and Moodle (VLE tools), while others used content management tools such as SharePoint. Others also used portal technologies or general institutional web-based tools. On the other hand, the German DFG encouraged and funded projects that developed new software following open source principles, while also funding projects that applied existing solutions (Carusi and Reimer, 2010: 46). In the USA, the focus has been more on the development of science gateways (or portal technology and gridware), as well as the creation of hubs (cloud driven tools). In the Netherlands, a flexible approach was followed, and funded projects were given the freedom to test any environment that would meet their needs. This can be seen from the various tools used in the projects, e.g. Liferay, SharePoint, RIC, Fedora, Zotero, Web 2.0 tools such as Google Apps and some Open Source tools.

Adaptations of shelf-ready tools like SharePoint and Sakai have been applied in a number of projects (e.g. PARTNER, Tales of the Revolt, RIC, etc.) across all four countries, necessitating outside help from suppliers to make things easier (Van der Vaart, 2010: 36).

3.3.3 Functional Aspects

Common to all the VRE programmes across the four countries, was a focus on collaboration and sharing of ideas between researchers. All four countries also focused on supporting the research lifecycle with their VRE programmes, and all four supported single, interdisciplinary and cross-institutional research. In the Dutch and German projects, data sharing was found to be an important and central feature, while the USA Science Gateways project also showed that "there is an increasing interest among researchers to make more data available." On the other hand, the UK programme showed the importance of evaluating/assessing the success of the selected VRE projects (Carusi and Reimer, 2010: 52-53). Van der Vaart (2010: 38) found that the use of collaboratories (VREs) led to the rethinking of research methods in the UK in JISC's projects, as well as in the Dutch projects, e.g. online text editing with categorisation and filtering of annotations. Finally, the Science Gateways project of the USA found VREs to be a useful interface to supercomputing resources (Carusi and Reimer, 2010: 52)

3.3.4 Policy / Legal / Financial Aspects

Three of the countries - the UK, the Netherlands and Germany each have a national institution that funds and drives the main VRE initiatives in their respective countries. In the USA, however, only the TeraGrid, a sub-project of the Science Gateways project, is funded by the National Science Foundation. The UK furthermore had a joint government-commercial venture between the British Library and the Microsoft Company in developing the RIC project.

Carusi and Reimer (2010: 32-33) found that a major challenge faced by all the VREs was that of sustainability. This would entail long-term support in terms of further funding, the development of business models to make VREs self-sustaining, as well as acceptance and use by the communities they are intended for. The Dutch SURFshare project furthermore highlighted the problematic issue of sharing resources, which require institutional subscriptions, with researchers at other institutions not having access to the subscriptions. The German eSciDoc project, in turn, stressed the importance of librarians in checking “that open access publications do not violate any third-party rights before publication” (Carusi and Reimer, 2010: 54, 73). Results from the UK myExperiment project showed that it is not always in the interest of individual researchers to share their data and scientific workflows until they are certain that they have obtained the complete value from them. This means that total open access to everything is not possible, but rather an approach to reserve some rights (Carusi and Reimer, 2010: 81).

3.3.5 Cultural Aspects

Comparisons between the countries showed that the UK is well advanced in its understanding of the VRE concept and has the world’s best structured programme of VRE developments so far (Carusi and Reimer, 2010: 12). The Netherlands focused more on the humanities and social sciences, while in the other countries, the focus was more mixed, in that it focused on different disciplines. The feedback from the German projects also showed the difficulty in building appropriate services and solutions that facilitate collaborations across discipline boundaries (Carusi and Reimer, 2010: 74). Nonetheless, trust was highlighted by the German project eSciDoc as a key factor in the uptake of a VRE, concerning the developers as well as the technical infrastructure (Carusi and Reimer, 2010: 74).

All four countries discussed focused on supporting the research cycle. The next section will discuss this concept further, as well as the various VRE components and tools that can be found in the different stages of a research cycle.

3.4 RESEARCH CYCLES AND VRE COMPONENTS

3.4.1 Research Cycles

The research process is often described in an abstract manner as a research cycle, to give clearer understanding of the various features of the process. Using these features as a foundation, one could get a clearer picture of what a generic VRE might look like (Voss and Procter, 2009: 178-179). The different VRE projects discussed above all addressed various aspects of the research cycle, and showed that a research cycle could contain a number of components. The RIC divided the Research Cycle into four main components, namely idea discovery and design; obtaining funding; experimenting, collaborating and analysing; and dissemination of findings (Barga, Andrews and Parastatides, 2007: 31). According to Voss and Procter (2009: 179), the research cycle has the following components:

- "The initial exploration of an idea and the acquisition of basic knowledge about a new research method like social simulation";
- Obtaining funding;
- Data acquisition and collection;
- Data analysis using computational resources;
- Storing intermediate results and outputs;
- Exchanging information and networking with other researchers; and
- Storing of final outputs on institutional and national repositories.

Humphrey (2006) calls the research cycle a Knowledge Transfer Cycle. He divides his cycle into two levels. The top level consists of data discovery, data repurposing and the production of new data. The bottom level of the cycle runs parallel to the top level and consists of the following components: Study concept and design; data collection; data processing; data access and dissemination; and analysis leading to research outcomes. He also proposed a second cycle, where he identified research communications during the research cycle. The components of this cycle are:

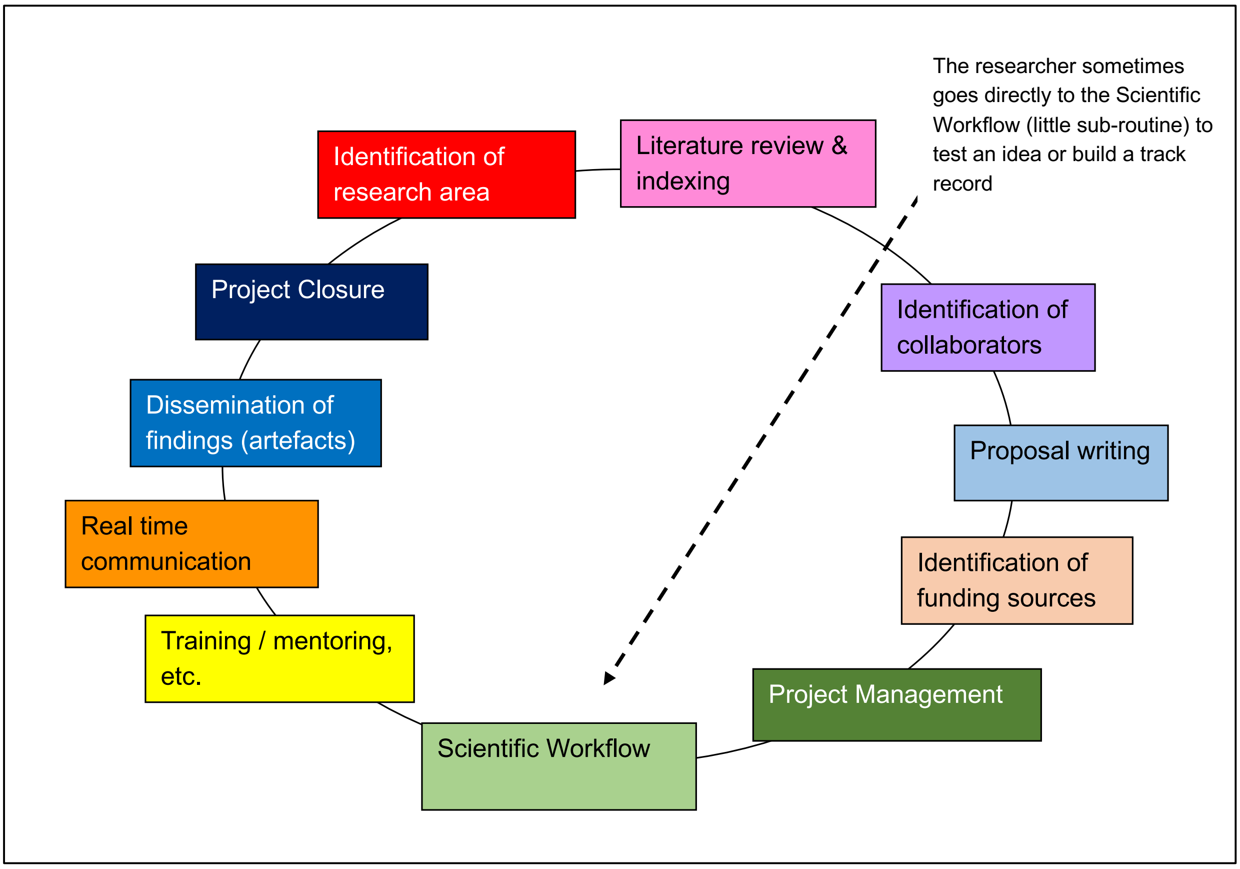
- Conceptualising (e.g. e-mails, letters, literature reviews);
- Initialising (e.g. grant applications, meeting minutes);
- Analysis (e.g. presentations, conferences and seminars);
- Initial results (e.g. grant reports, technical reports, thesis);
- Formalising (e.g. journal articles, books, curricula content, policy); and
- Popularising (e.g. popular literature, newspapers, practice).

Pienaar and Van Deventer (2009) compiled a comprehensive research cycle in their research on the need of a VRE for South African Malaria researchers. They identified the following stages in their research cycle:

- Identification of research area, which is comparable to Voss and Procter's (2009: 179) initial exploration of an idea and Humphrey's (2006) conceptualising;
- Literature review and indexing, which is comparable to Voss and Procter's acquisition of basic knowledge and Humphrey's (2006) conceptualising;
- Identification of collaborators, which is comparable to Humphrey's (2006) initialising;
- Proposal writing, which is comparable to Humphrey's (2006) initialising;
- Identification of funding sources, which is comparable to Voss and Procter's (2009, 179) obtaining of funding and to Humphrey's (2006) initialising;
- Project management;
- Scientific workflow, which is comparable to one of RIC's four stages in the research cycle (see 3.2.1), namely experimenting, collaborating and analysis. It also encompasses Voss and Procter's (2009: 179) data acquisition and collection, data analysis using computational resources, and storing intermediate results and outputs, as well as Humphrey's (2006) analysis and initial results;
- Training and mentoring;
- Real time communication, which is comparable to Voss and Procter's (2009: 179) exchanging of information and networking with other researchers;
- Dissemination of findings (artefacts), which connects with Voss and Procter's (2009: 179) storing of final outputs on institutional and national repositories, and Humphrey's (2006) formalising and popularising.

Pienaar and Van Deventer's (2009) cycle will be used as a basis to identify possible components for a VRE.

Figure 3.3: The Research Cycle as identified by Pienaar and Van Deventer (2009) and Van Deventer et al. (2009), and Van Deventer (2015)

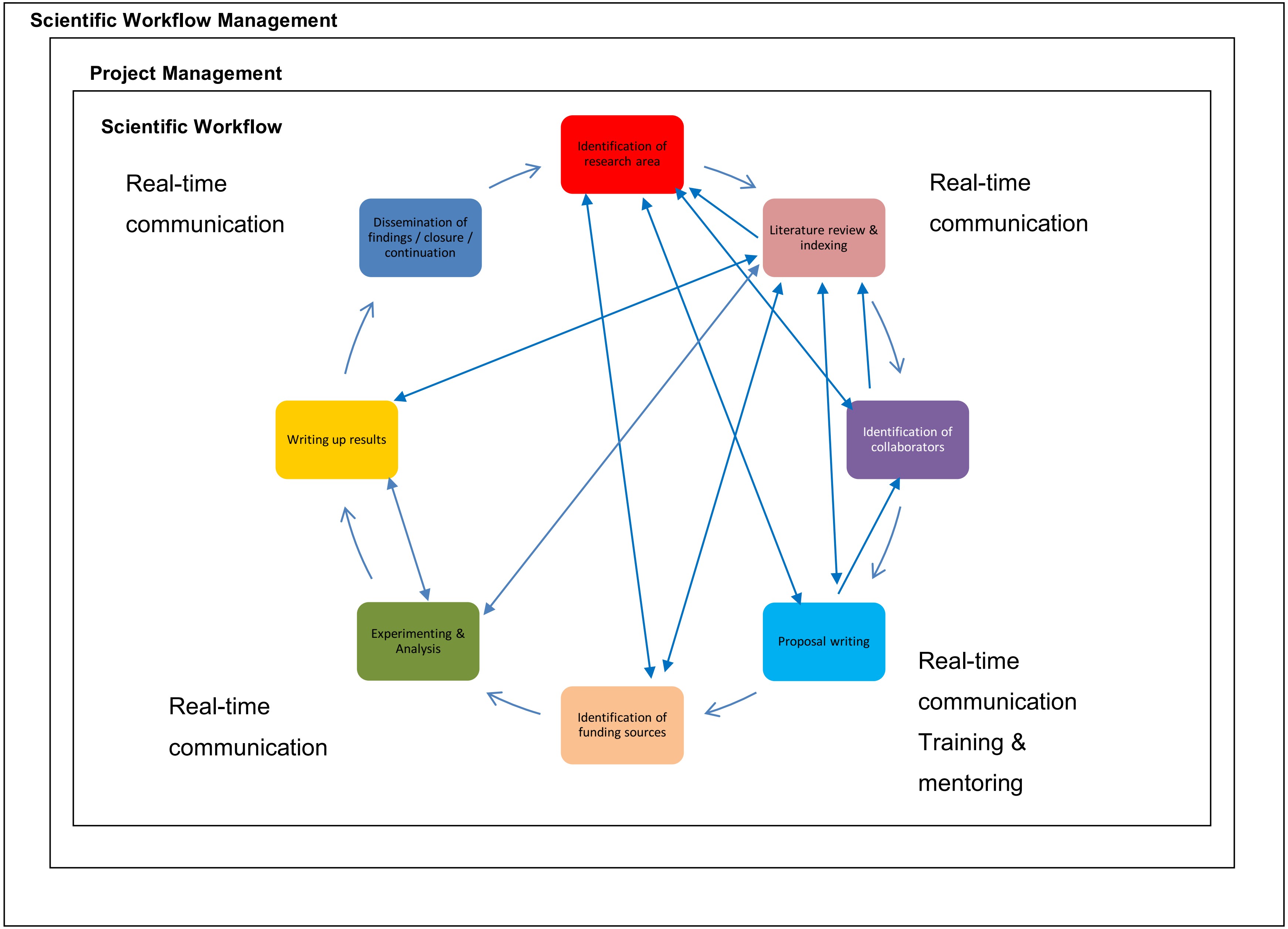

The researcher sometimes goes directly to the Scientific Workflow (little sub-routine) to test an idea or build a track record
Identification of research area
Literature review & indexing
Identification of collaborators
Proposal writing
Identification of funding sources
Project Management
Scientific Workflow
Training / mentoring, etc.
Real time communication
Dissemination of findings (artefacts)
Project Closure

Pienaar and Van Deventer (2009) acknowledge that research is an iterative process rather than a definitive cycle, but nevertheless still presented the research process as a research cycle with the different stages as shown in Figure 3.3. This researcher however does not agree with the model as presented by Pienaar and Van Deventer (2009), Van Deventer et al. (2009) and Van Deventer (2015), and would rather adapt the model by substituting scientific workflow with scientific experimentation and analysis, and use the concept 'scientific workflow' as encompassing the whole research cycle. The concept of RDM that was described by Pienaar and Van Deventer (2009) as part of the scientific workflow component, is seen as a process that takes place in each of the components of the research cycle. The flow of RDM in the research cycle is more elaborated upon in Chapter 5.

This researcher regards the management of the scientific workflow as an action that can be done through a process of project management. Furthermore, the writing up of the results, even though implied, is an important component that is missing from Pienaar and Van Deventer's (2009) model, and should be added in an adapted model, while training and mentoring as well as real-time communication are components that take place throughout all the stages of the research cycle, not just at specific points in the research cycle. In Van Deventer (2015), a closure stage was added just after the dissemination of findings stage. The researcher of this study agrees that there is a closure stage, but views the dissemination stage as the closure stage, although not in all instances. In some instances, the research lifecycle could be continuous, where-as a research project continues indefinitely.

The adaptation of Pienaar and Van Deventer (2009), Van Deventer et al. (2009), and Van Deventer's (2015) model can be seen in Figure 3.4. A researcher can identify a research area and then do a literature review, or first do a literature review and then identify a research area. Following this, the researcher can identify possible collaborators, but can also first write a proposal and then identify possible collaborators. During the proposal writing, the researcher might have to do a further literature study before continuing with the research process. While writing the proposal, the identification of the research area might change. The identification of the research area can furthermore lead to the identification of funding resources, but this can also be the opposite, where the identification of funding resources can lead to the identification of a research area. In the experimenting and analysis stage, a further literature study might be necessary before writing up the results.

Figure 3.4: Adapted Version of Pienaar and Van Deventer (2009), and Van Deventer et al.'s (2009) Research Cycle



The following section will focus on the various VRE components of a research cycle.

3.4.2 VRE Components

Myhill, Shoebridge and Snook (2009: 230) identify common components of a VRE that compare well with the stages in a research cycle as proposed by Voss and Procter (2009: 179), Humphrey (2006) and Pienaar and Van Deventer (2009). In other words, the stages in itself could also be seen as components of a VRE. These components, according to Myhill, Shoebridge and Snook (2009: 230) are:

- Identifying a research project;
- Identifying funding streams;
- Identifying project partners;
- Collaborating on a research proposal;
- Managing the project, including expenditure and grant compliance;
- Collaborating on research information;
- Writing research reports and other outputs; and
- Disseminating results.

Voss and Procter (2009: 179) list possible VRE processes that could be translated into the following VRE components:

- Authentication component (“authenticate using an authentication service” and “find out who has access to a resource and what they can do with it”);
- Communication component (“communicate and collaborate with colleagues”);
- Data transfer component (“transfer data”);
- Literature review and indexing component (“configure a resource”);
- Computational component (”invoke a computation”);
- Data citation component (“re-use data and give credit to the original producer”);
- Data repository component (“archive output data and runtime data”);
- Dissemination of findings component (“publish outputs, informally through blogs or wikis and formally through conference or journal papers”);
- A search function component (“discover what resources are available”);
  - Scientific workflow component (“monitor the state of a resource or process”);

- Identification of collaborators component (“maintain awareness of who is currently doing what”); and
- Intellectual property management component (“find out where particular data has come from and how it was processed”).

The University of London Library, according to Chambers (2002: 389-390), proposed the following components for developing a VRE model:

- Virtual research library support;
- Research-related information;
- Online secure research repository (OSRR);
- Online research support mechanisms;
- Tracking of research activity and achievement;
- Research output repository;
- Software evaluation; and
- Researcher involvement.

Klyne (2006) identified a list of requirements for VREs, which could be translated as VRE components. These include:

- Access to best-practice documentation, and support for best practices, within the VRE;
- Support for researchers' day-to-day activities;
- Capturing and storing of collaborative discussions;
- Access to searchable databases that have digital (digitised) artefacts;
- Providing support for the training of new researchers;
- Searchable list of lectures, conferences and other events;
- Capability to locate other researchers;
- Selective distribution of information;
- Support for grant applications;
- Provision of spaces and forums where internal communication and recruitment can take place; and
- Access to HPC facilities where modelling can be done.

Sergeant, Andrews and Farquhar (2006) list a number of VRE components as derived from the EVIE VRE project. These include:

- Finding and acquiring published information, for example journal articles, conference proceedings and literature;
- Collaboration with associates at the university and at other organisations or tertiary institutions;
- Finding information about funding opportunities, application for funding and the management of funded projects;
- Sharing or archiving of research results, e.g. data sets, technical reports, preprints, post-prints, etc.

Di Muro and Saunders (2008), on the other hand, identify four core components that should comprise a VRE. These are:

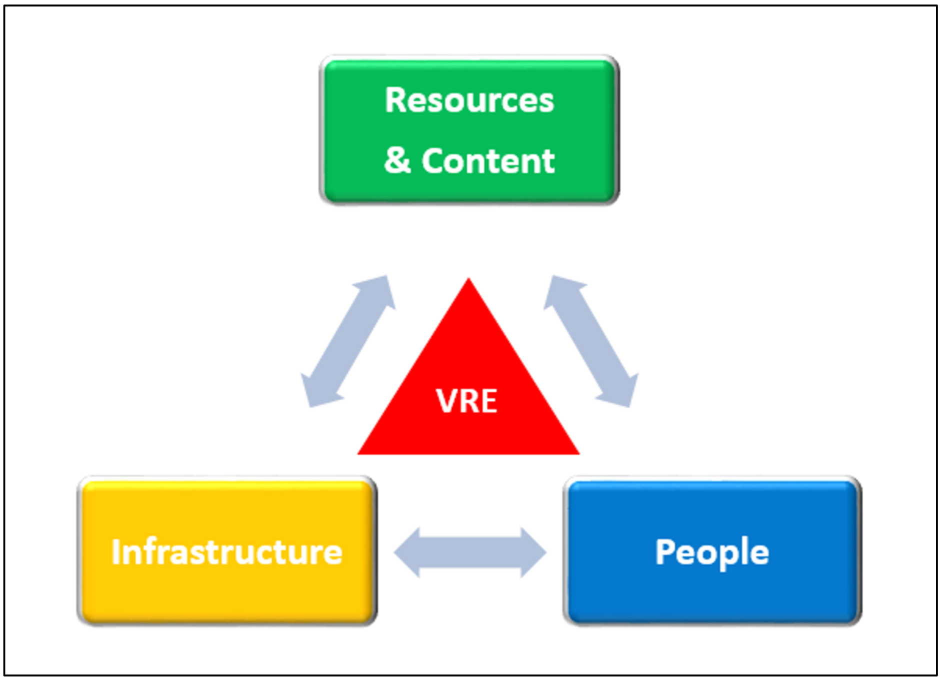
- Collaboration (the central function that induce communication and networking in communities of enquiry);
- Knowledge (gives access to scholarly information, that is: “access to library resources”, databases, open access repositories and other academic documents);
- Data (this gives “access to raw experimental and statistical data sets” as well as “the tools to analyse them”); and
- Experimentation (the most discipline-specific and distinctive element of a VRE, and can include tools with huge “processing power to conduct simulated experiments”).

Keraminiyage, Amaratunga, and Haigh (2009b: 133-134) highlight the human component of a VRE by listing four success factors of research collaborations:

- Trust, commitment, ability and leadership;
- Transparency and clarity;
- Communication; and
- Monitoring.

Frederique van Till from JISC describes a VRE as a platform with three components: resources and content as one component, and infrastructure and people as the other two components (Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London).

Figure 3.5: VRE Platform With 3 Components



During the development of VREs, according to Van Till, people use different approaches by emphasizing one of these components. In some VREs, demands of the people will determine which resources and which infrastructure will be needed. In other VRE developments, the resources and content that are available will determine which audience (people) will find this useful, and which infrastructure might be needed. In other VRE developments, the infrastructure is a given and people are told how to use it, and what content and resources they can use/place in it.

By combining and integrating these components as identified by Chambers (2002: 389-390), Di Muro and Saunders (2008), Keraminiyage, Amaratunga and Haigh (2009b: 133-134), Klyne (2006), Myhill, Shoebridge and Snook (2009: 230), Pienaar and Van Deventer (2009), Sergeant, Andrews and Farquhar (2006), Van Deventer et al. (2009), Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London, and Voss and Procter (2009: 175, 179), a list of possible VRE components relevant to this study can be compiled and then matched to the stages of the research cycle as presented in Figure 3.4, to ensure that a VRE successfully support the research process. Possible tools (See Addendum A for a list) for the different components are also listed.

Below in Table 3.2 is a possible matching of these components and tools to the different stages of the research cycle.

Table 3.2: Matching Of VRE Components And Tools To The Research Cycle

STAGE	POSSIBLE VRE COMPONENTS	POSSIBLE VRE TOOLS
IDENTIFICATION OF RESEARCH AREA	• Personal networks (Human component, communicate and collaborate with colleagues, maintain awareness of who is currently doing what). • Hypothesis Formulation. • Literature search (discover what resources are available, research-related information, tracking of research activity and achievement). • Funders (research related information).	• Searching tools, for example web search engines such as Google, and federated library search engines, e.g. WorldCat by OCLC. • Wikis, blogs, portals. • Scholarly Databases. • LinkedIn. • RSS Feeds. • Open Access Repositories using software such as DSpace or Fedora.
LITERATURE REVIEW AND INDEXING	• Search function (discover what resources and knowledge are available) (data discovery/collection). • Referencing.	• Web search engines, e.g. Google Scholar. • Scholarly Databases. • Reference databases, e.g. RefWorks, Endnote, Mendeley. • Internal shared database of indexed articles.
IDENTIFICATION OF COLLABORATORS	• Personal networks (Human component including issues of trust, of who will take leadership, and transparency and clarity, communication and collaboration with colleagues, and awareness of who is currently doing what).	• Search Engines, e.g. Google. • Expertise lists, e.g. Research Africa. • Social tools e.g. LinkedIn, ResearchGate, and Flickr, which has interest groups. • Citation databases such as ISI (Web of Knowledge) and SciVerse Scopus.
PROPOSAL WRITING	• Word processing. • Document management.	• Word Processing tools, e.g. MS Word, Google Docs. • Collaboration tools such as Skype and Dropbox, wikis,

		Evernote, Google + Hangouts. • Job submission tools (e.g. Sakai). • Document Management Systems, e.g. SharePoint.
IDENTIFICATION OF FUNDING SOURCES	• Identify funders/funding.	• A website with a list of funders easily accessible. • E-mail alerts. • RSS Feeds.
EXPERIMENTING AND ANALYSIS	• High-Performance computation (invoke a computation). • Management of intermediate research results, using data analysis software. • Experimentation. • Simulation. • Visualisation. • Validation. • Data storage	• High-performance computation resources, Kepler, 'R', Taverna, Triana. • In silico experimentation software, which can include simulation software, modelling software, e.g. 'R', Taverna, JChem chemical structure tool. • Statistical analysis tools e.g. 'R'. • General or domain-specific analytic and visualization software, e.g. 'R', TextGrid, Triana. • Data analysis software, e.g. Archer e-Research toolset, Kepler, 'R', ScratchPads, Taverna, TextGrid, Triana. • Electronic Lab Book, e.g. Open WetWare, and eCAT. • Linked storage which can consist of Science / Research clouds as well as cloud services, e.g. Google Drive and Dropbox. • Document management systems, e.g. Alfresco can also be used to manage and store data short term.
WRITING UP RESULTS	• Word processing. • Using spreadsheets. • Using presentation software. • Using social media. • Using data citation tools	• Word Processing tools, e.g. MS Word, Google Drive. • Spreadsheet tools, e.g. MS Excel. • Presentation software such as MS PowerPoint and Prezi. • Blogs and wikis. • Data citation tools, e.g. Mendeley, and Endnote

DISSEMINATION/OUTPUT OF FINDINGS (artefacts)	• Peer Review. • Publishing outputs, informally through blogs or wikis and formally through conference or journal papers. • Archiving (online research output repository). • Long-term Preservation and Management of research results through data curation and Management (archive output data and runtime data).	• Do closed or open peer review through tools such as Skype, Google Drive, e-mail, Dropbox, and ISI. • Publish outputs formally in journals, books, reports etc. (This could be in subscription based publications or in open access). • Publish outputs informally through blogs, wikis, Flickr, and Slideshare. • Access to citation databases e.g. ISI and SciVerse Scopus to determine best publication. • Archive/disseminate/publish research results through repositories, e.g. FedoraCommons, D-Space, Figshare, Github, Liferay, MS SharePoint, OpenWet-Ware, ScratchPads, and TextGrid. • Data curation and management tools, e.g. data repositories (DSpace, eCAT, e-SciDoc, FedoraCommons, Figshare, Github, HUBzero, Kepler, myExperiment, NanoHUB, OpenWetWare, RIC, ScratchPads, TextGrid, and Triana. • Disseminate results through webinars & virtual conferencing tools, e.g. Skype, Google Talk, Google Hangouts etc.
Researcher involvement is a component that has to be sustained throughout the research cycle		

The following components encompass the research cycle, in other words take place throughout the research cycle and have been matched with possible components/tools that will enhance their effectiveness:

Project Management:

- Job submission tools (e.g. Sakai has a job submission tool);
- Project Management Systems, e.g. MS Project.

Training/Mentoring:

- E-learning tools: e.g. E-learning systems for researchers, using tools such as Sakai, Moodle or Blackboard.

Real-time Communication:

- Collaborative Interfaces, e.g. Academia.edu, Alfresco, Archer e-Research toolset, Blackboard, Chisimba, Drupal, eCAT, e-SciDoc, Evernote, Flickr, Google Apps for Education, Google Drive, Google Talk, Google+ Hangouts, HUBzero, Kepler, Liferay, Mendeley, Moodle, MS SharePoint, My Experiment, nanoHUB, OpenWetWare, ResearchGate, RIC, Sakai, ScratchPads, Skype, Slideshare, Taverna, TextGrid, Triana, uPortal, and ZEENOV Agora, Zotero.
- Instant Messaging (IM) tools, e.g. Blackboard, which has an IM and SMS facility, and Whatsapp, a mobile tool.
- Announcements, e.g. Academia.edu, Blackboard, Chisimba, Google+ Hangouts, Moodle, myExperiment, ResearchGate, RIC, Sakai, ScratchPads, Slideshare, and ZEENOV Agora.
- Audio conferencing, using tools such as Google Talk.
- Video conferencing, using tools such as Blackboard, Sakai, Skype, Google Hangouts and ZEENOV Agora.

Scientific Workflow Management:

- Reference Management, using reference management tools such as Refworks, Endnote and Mendeley.
- Workflow management (transfer data, re-use data and give credit to the original producer, monitor the state of a resource or process, online research support mechanisms), using specific workflow management tools, e.g. myExperiment, Taverna and Kepler.

By using insights gained from the study of the VRE programmes in the four countries, as discussed earlier, as well as the matching of components and tools to the research cycle, a possible conceptual VRE framework can be compiled.

3.5 POSSIBLE CONCEPTUAL VRE FRAMEWORK

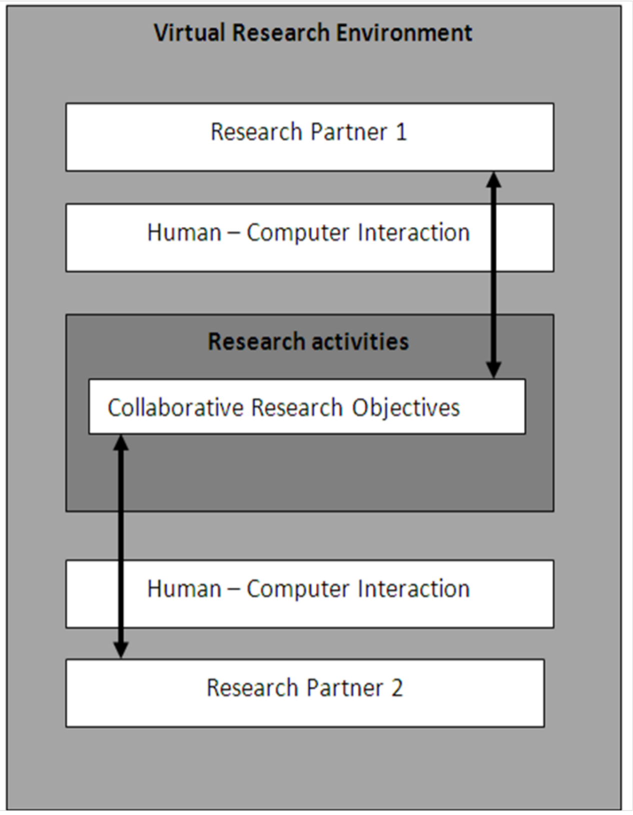
In the discussion of a possible conceptual VRE Framework, it is important to have a clear understanding of what is meant by ‘conceptual framework’. Weaver-Hart (1988: 11) describes a conceptual framework as “a structure for organising and supporting ideas; a mechanism for systematically arranging abstractions; sometimes revolutionary or original, and usually rigid.” In other words, it can offer a “theoretical overview of intended research as well as order within that process” (Leshem and Trafford, 2007: 96). Rudestam and Newton (1992: 6), on the other hand depict a conceptual framework as “simply a less developed form of a theory,” comprising of statements that link abstract concepts to empirical data. Theories and conceptual frameworks according to Rudestam and Newton (1992: 6) “are developed to account for or describe abstract phenomena that occur under similar conditions.” Conceptual frameworks, in other words, can be used to connect theory with practice. Robson (1993: 150-151) emphasizes conceptualisation as meaning-making in research by stating that the development of a conceptual framework forces one to be explicit about what you think you are doing. According to Robson (1993: 150-151) it also “helps you to be selective; to decide which are the important features; which relationships are likely to be of importance or meaning; and hence, what data you are going to collect and analyse.” Maxwell (1996: 25, 37) describes a conceptual framework in terms of a concept map, which is “a visual display of your current working theory” or “a picture of what you think is going on with the phenomenon you’re studying.”

A review of literature shows various kinds of conceptual frameworks focusing on e-Research, e-Research infrastructure, subject specific VRE architecture, web-based support systems, etc. The researcher chose the following examples of conceptual frameworks from literature as they represent an array of frameworks starting from the most simplistic to the most complex.

3.5.1 Keraminyage, Amaratunga And Haigh's (2009b: 129-142) Visualised Structure Of A VRE

Keraminiyage, Amaratunga and Haigh (2009b: 131-132) visualise the structure of a typical VRE in Figure 3.6. In this visualisation, they integrate various research partners that are frequently geographically separated through a human-computer interface, but are striving to attain a shared set of research objectives by concluding a range of research activities (Keraminiyage, Amaratunga and Haigh, 2009b: 132).

Figure 3.6: The structure of a typical VRE (Keraminiyage, Amaratunga and Haigh, 2009b: 131)


Virtual Research Environment
Research Partner 1
Human – Computer Interaction
Research activities
Collaborative Research Objectives
Human – Computer Interaction
Research Partner 2

The partners are connected by the VRE to the research objectives via the human-computer interface. The VRE also aid them in realizing the shared objectives of the research collaboration. They identify two main elements of a VRE: the human-computer interface, as well as the functionalities embedded to attain success factors of collaborative research (Keraminiyage, Amaratunga and Haigh, 2009b: 132).

Keraminiyage, Amaratunga and Haigh (2009b: 133) list a number of success factors for collaborative research as found in Barnes, Pashby and Gibbons (2006: 397-398): universal success factors such as “mutual trust, commitment, good personal relationships, continuity, flexibility, and leadership”; project management success factors such as “clearly defined objectives, clearly defined responsibilities, a mutually agreed project plan, realistic aims, adequate resources, defined project milestones, a simple collaborative agreement, regular progress monitoring, effective communication, etc.” They also cite Dodgson’s (1996) discussion on trust in research collaborations, where he identifies three types of trust, namely contractual trust, “where all the parties trust that each of the parties will adhere to agreements and promises,” competence trust, which assures the “abilities of partners to each other,” as well as goodwill trust, which creates “mutual respect for each other, respectively” (Keraminiyage, Amaratunga and Haigh, 2009b: 133).

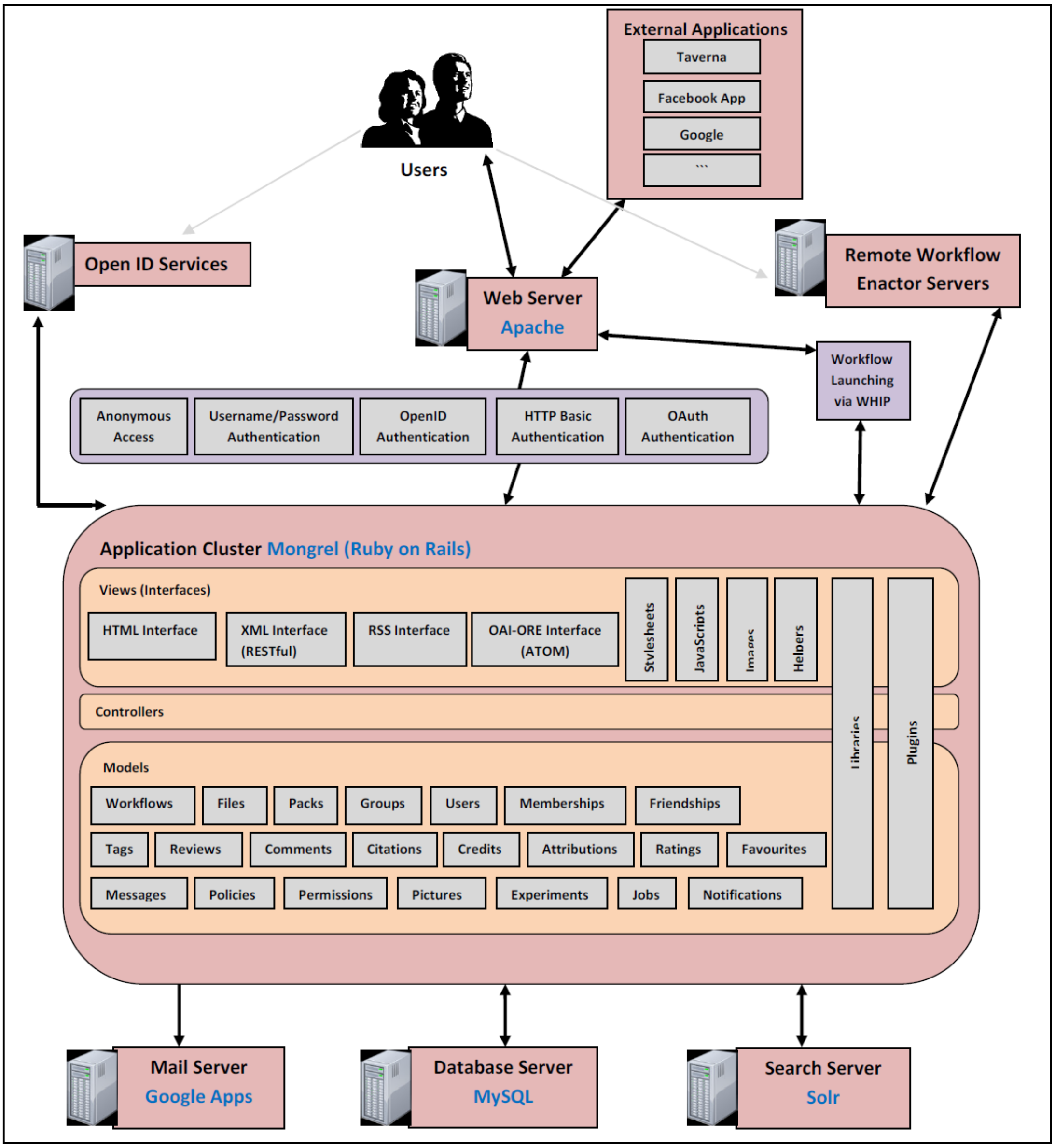
Keraminiyage, Amaratunga and Haigh’s (2009b) structure, however, does not give a detailed conceptual framework of a VRE. It only provides a broad overview. They do mention the human component of a VRE, but do not elaborate on who the various research partners are; similarly, they do mention a human-computer interface, but nothing is said about the various hardware or software components needed, or standards, protocols or specifications. The aspect of collaboration is mentioned though, as well as the fact that the research partners have a shared set of research objectives and that they conclude a range of research activities. The research activities or research cycle conversely is not elaborated upon. Their structure nevertheless touches on some important components that can be of value in the design of a conceptual framework. The success factors for collaborative research, such as mutual trust, commitment, good personal relationships, continuity, flexibility, leadership, clearly defined objectives, clearly defined responsibilities, a mutually agreed project plan, realistic aims, adequate resources, defined project milestones, a simple collaborative agreement, regular progress monitoring, and effective communication can be of great value in a VRE conceptual framework, and has been added to this researcher’s design of a possible conceptual framework of a VRE.

3.5.2 De Roure et al.’s (2009) Illustration of the myExperiment Architecture

myExperiment was created along the lines of an interpretation of Web 2.0 design principles, in other words, it has an open environment capability, within the context of a VRE (De Roure et al., 2009). Figure 3.7 shows the architecture of one instance of myExperiment as illustrated by De Roure et al. (2009). In this instance, the service is hosted on two servers: a web frontend, consisting of an Apache web server, and a database backend consisting of a cluster of Ruby on Rails (web application framework) processes operating on distinct ports utilising the Mongrel Cluster software. Ruby on Rails makes it possible to leverage numerous resources to build features for users speedily (De Roure, et al., 2008). Authentication is done via external OpenID Services or by using the internal username/password mechanism (De Roure, et al., 2008). To authenticate API access, myExperiment utilises the OAuth protocol. For example, the OAuth protocol can be used to authenticate that a user has given a service consumer access to a service provider. It is also an exclusive key that has specific privileges allocated to it. Using OAuth, several keys can be created to be used with one service, but each having diverse privileges (De Roure, et al., 2008). HTTP protocol is used to access all the interfaces to the myExperiment functionality. The Apache web server provides an HTML based web interface to end users. The interfaces within the Application Cluster are accessed by means of a web server that manages “load balancing over a cluster of mongrel application servers” (De Roure et al., 2008). A mechanism is automatically provided to REST access through the Ruby on Rails framework, but the designers of myExperiment decided to manage the REST API separately so that they could react more easily to the needs of API users. This REST API is operated through an XML specification, loaded and edited in Microsoft Excel (De Roure et al., 2008). This makes it possible to “create an independent API specification” with the added advantage that it is not spread out across many model files, but is situated in one place (De Roure et al., 2008). External applications (e.g. Facebook, Taverna, and Google etc.) also have the ability to access the other interfaces, especially the managed RESTful API (De Roure et al., 2009). Elements of the data model comprise: workflows, files, packs, users, groups, memberships, friendships, tags, reviews, comments, citations, credits, attributions, ratings, favourites, messages, policies, permissions, pictures, experiments, jobs and notifications.

The mail server, database server, and search server at the bottom layer of the figure, as well as the Remote Workflow enactors in the second layer are all separate and external systems to which the main application cluster connects (De Roure et al., 2008). "The database which is the major component of the Ruby on Rails system, is hosted on the second server in the form of MySQL." This server houses the Solr search server, that is described by De Roure et al. (2008) as a "Java implementation of the Lucene search library" operating as a Java servlet in Tomcat. External Applications (APIs) can also be developed and plugged into the system, e.g. Taverna workflow system, Facebook Apps, and Google Gadgets, etc.

Figure 3.7: The implementation architecture of a MyExperiment Server Instance (Adapted from De Roure et al., 2009)



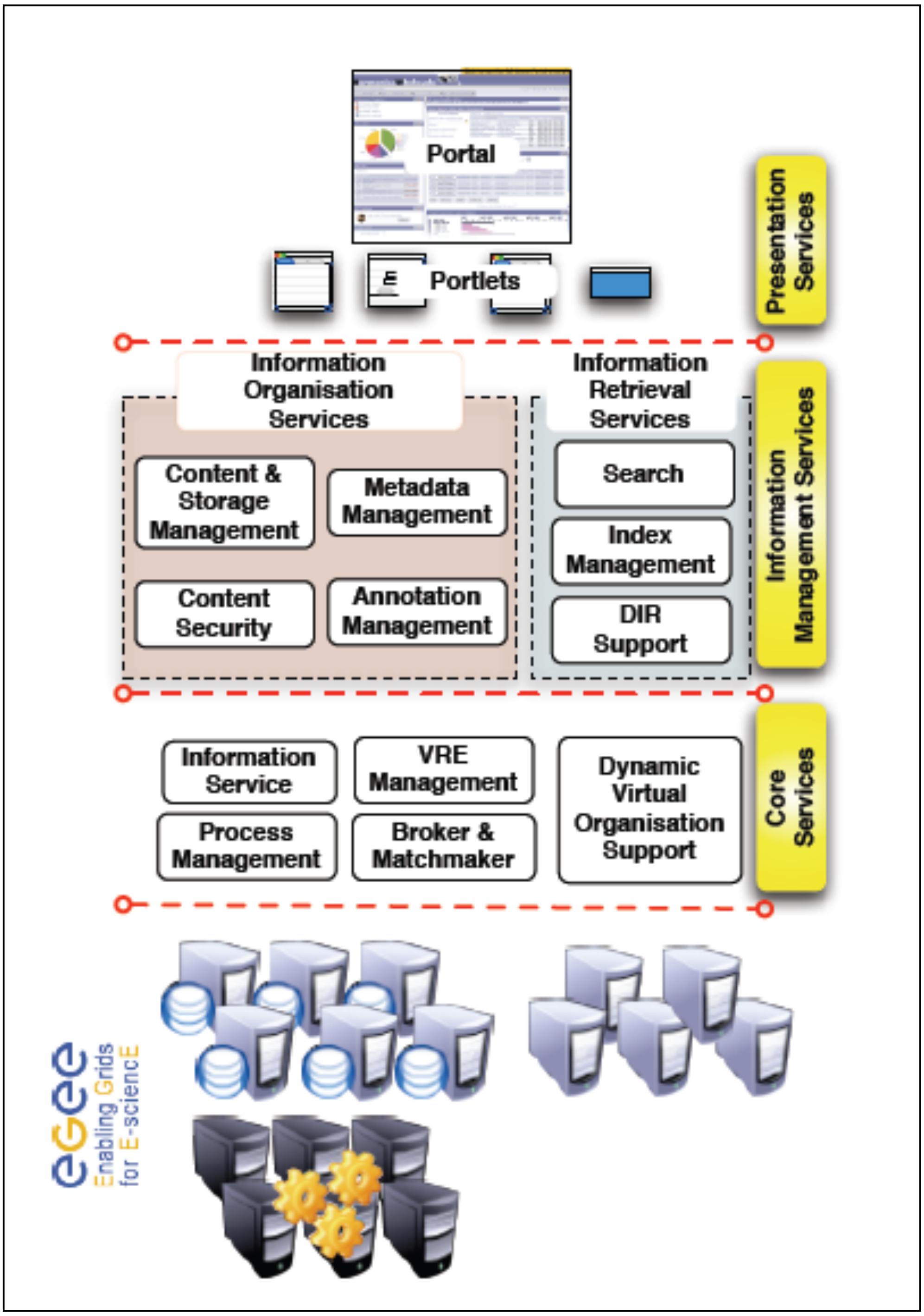
The illustration of the myExperiment architecture by De Roure et al. (2009) in Figure 3.7 does not fully cover all the aspects of a conceptual framework for a VRE. It does mention the user and the authentication tools that give each user different privileges and access, but it still does not specify the different types of users that will use a VRE (i.e. the human components), or mention anything about the collaboration between them. It also does not say much about the research process or cycle. It furthermore doesn't mention any success factors (called important policy components by this researcher in Figure 3.12b later in this study) for collaborative research as in Keraminyage et al.'s (2009b) structure of a VRE. Another component not mentioned is the possible hardware that users of the system will use, e.g. PC, tablets, cell phones, digital cameras, etc.

The valuable contributions of this illustration are the detailed description of its authentication mechanisms, the protocols mentioned, the various interfaces in the applications cluster, and the possibility of plugging in external APIs. It does mention some useful components in its models layer, such as tags, files, workflows, reviews, comments, citations, credits, attributions, ratings, favourites, messages, policies, permissions, pictures, experiments, jobs, notifications; however, the following, which could potentially be external APIs, are not mentioned: word processing, referencing, document management, computation, electronic lab books, simulation, visualisation, data analysis, data curation, project management, publishing, e-learning, repositories, intellectual property management, etc.

3.5.3 Simeoni et al.'s (2008) Illustration Of gCube Architecture / Framework

gCube, a software framework designed with the intention to construct e-infrastructures supporting VREs, was started in 2004, and was managed by an international consortium, partly funded by the European Community (Candela, Castelli and Pagano, 2010, 32; Simeoni et al., 2008: 1). Its core functionality was designed "and initially used in testbed infrastructures for the Environmental Monitoring and the Cultural Heritage domains" (Simeoni et al., 2008: 1). gCube had a rich assortment of mediator services, which were used to interface with existing infrastructure enabling technologies comprising cloud (e.g. Hadoop), grid (e.g. gLiote/EGEE) and data source (e.g. OAI-PMH) oriented approaches. These mediator services were used to unite the processing facilities, storage facilities and data resources of the external infrastructure conceptually, in order to become gCube tools (Candela, Castelli and Pagano, 2010: 32).

Figure 3.8: gCube Architecture (Simeoni et al., 2008)



According to Simeoni et al. (2008: 1), gCube consisted of approximately 140 components functionally spread out across three layers (illustrated in Figure 3.8). The top-layer consisted of a series of presentation services (portlets giving access to portals), which were used by VRE end-users and administrators to interact with services in the other layers. An application support layer was later added, between the Presentation Services and the middle layer (Frosini, 2016). The middle layer grouped services for information management into two groups. The one group consisted primarily of a stack of content-management services under the heading of Information Organisation Services. This group contained content and storage management (including storage replication and distribution services), content security, metadata management and annotation management (specialising in "the semantics of relationships in order to collate, describe, annotate, and transform cross-media content") (Simeoni et al., 2008: 1). The second group consisted of a runtime framework of search-management services, under the heading of Information Retrieval Services. This grouping processed and optimised structured and unstructured queries over a federation of advanced, "geo-spatial, or inverted indices of dynamically" chosen content resources (Simeoni et al., 2008: 1). The group contained the following components: a search framework, an index management framework, and distributed information retrieval (DIR) support (Simeoni et al, 2008: 2). A personalisation service was later added to this group (Frosini, 2016).

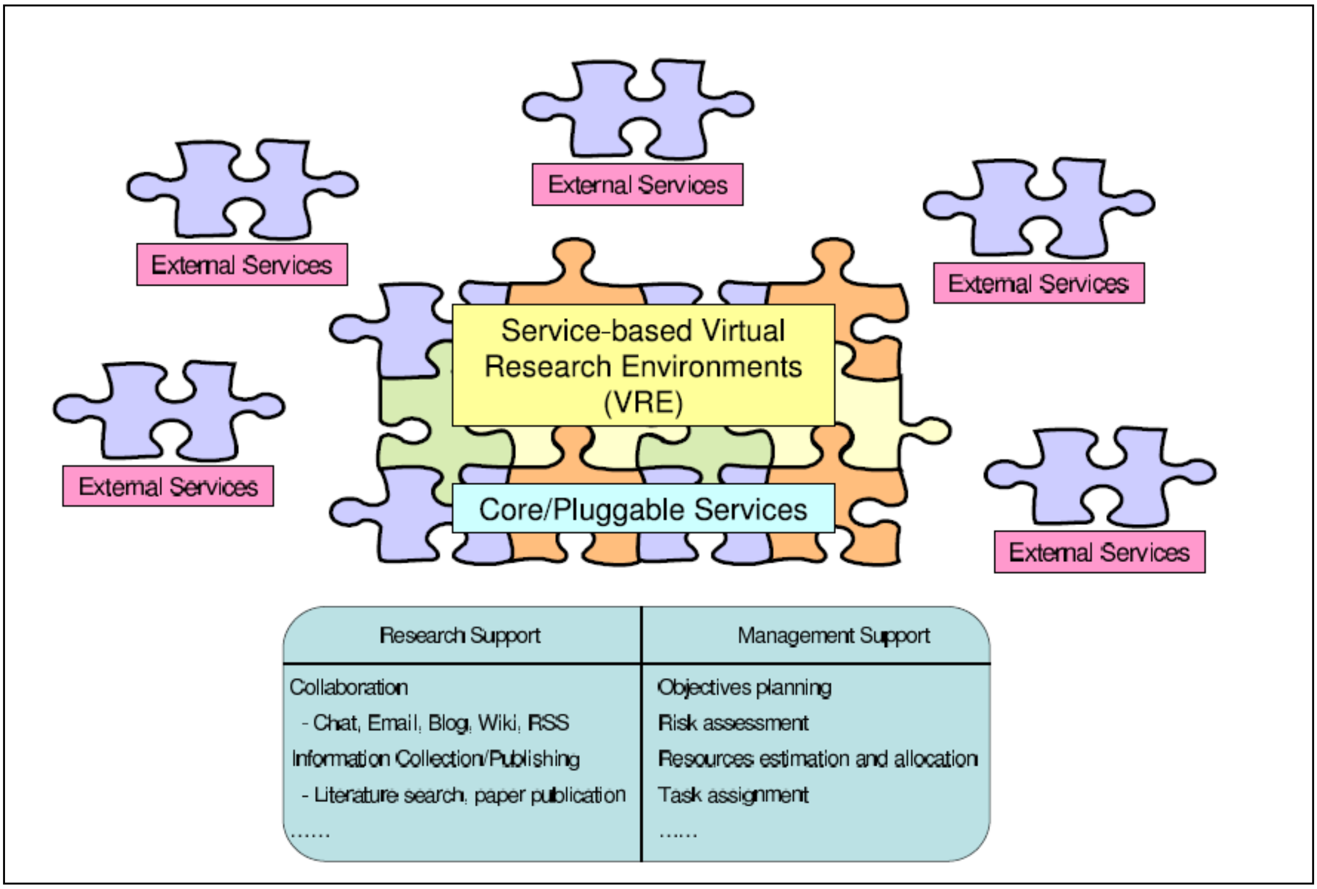
The bottom or Core Services layer contained "the services which confer autonomic behaviour to the whole system (Simeoni et al., 2008: 1). In this layer, the process management services utilised "graphically defined workflows of service invocations," distributing the enhancement, monitoring, and execution of individual stages across the infrastructure (Simeoni et al., 2008: 1). Users and services of the VRE were authenticated by the security services in the middle layer. The VRE management services component hosted service implementations and translated "interactive VRE definitions into declarative specifications for their deployment and runtime maintenance," and in particular, the initiation and dynamic configuration of the workflows that oversee their execution (Simeoni et al., 2008: 1). The brokering and match-making algorithms component controlled the deployment strategies that were "based upon static and dynamic information about the available hardware, data, and services" (Simeoni et al., 2008: 1). Information were gathered by a peer-to-peer network of information services (another component) and made available to all the services within the infrastructure. The information services component was later renamed information system; the dynamic Virtual Organisation support were renamed Virtual Organisation management; and process optimisation was added as another component, in a later version of the model (Frosini, 2016). The updated version of the model also had a gCube Container and a gCore Framework (Frosini, 2016).

Simeoni et al.'s (2008) depiction of gCube's architecture has valuable components that should be a in a VRE, but it does not cover all aspects that should be in a conceptual framework of a VRE. Their description does not say much about the users (researchers, the VRE facilitator, funders, developers, librarians, peer reviewers, etc). The research process (cycle) is also not mentioned. The idea of a presentation layer is valuable though, and has been built into this researcher's possible conceptual framework in the form of the Interface/platform layer (see Figure 3.12a later in this chapter). Futhermore, although Simeoni et al.'s (2008) illustration focuses a lot on content management and on searching of information, with related issues on storage, content security, metadata management and annotation management and information retrieval, nothing is mentioned about publishing, peer review, data management, invoking a computation, experiments, visualisation, collaboration and communication between users etc. On the other hand, valuable components that Simeoni et al.'s (2008: 2) illustration contributes are the VRE Management Services Component, as well as the brokering and match-making algorithms component, which controlled the deployment strategies on static and dynamic information about the available hardware, data, and services.

3.5.4 Yang And Allan's (2007) Service-Orientated Architecture (SOA) For VRE Systems

In Figure 3.9, Yang and Allan (2007: 540) describe a VRE system as having core pluggable services that meet the stated and tacit user requirements, for example an authentication and authorisation service, and a communication service. In addition, according to Yang and Allan (2007: 541), a VRE system should be expandable by plugging in new services or external services, so that new requirements can be met. They also see VRE systems as generally following a three-tier architecture (not shown in Figure 3.9), where web portals act as the presentation layer, with business logic and data layers behind it.

Figure 3.9: Yang And Allan's (2007) Service-Orientated Architecture



Research Support	Management Support
Collaboration - Chat, Email, Blog, Wiki, RSS Information Collection/Publishing - Literature search, paper publication	Objectives planning Risk assessment Resources estimation and allocation Task assignment ......

Web portals afford end-users with a single point of access to a variety of resources inside or outside the VRE (Yang and Allan, 2007: 541). These web portals are standards-based and Yang and Allan (2007: 541) identify two Java type portal standards: Java Portlet Specification 1.0, also known as JSR 168, and Web Services Remote Portlets (WSRP). Yang and Allan (2007: 541) then describe how Sakai, an open source e-learning system, matches the above mentioned VRE framework. According to them, Sakai can be divided into two sections: the Sakai framework and Sakai tools. The framework offers presentation and commons services to form a core system, whereas tools are pluggable and can be utilised for specific purposes, for example chat rooms and discussion, or used for common services such as retrieving current user information (Yang and Allan, 2007: 541). In their discussion on extensions to Sakai, Yang and Allan (2007: 541) stress the importance of portlets as web components that can be composed in web pages and portals. They also discuss how "the JSR 168 specification standardises communication between a portlet and its container, which enables development of re-usable portlets" (Yang and Allan, 2007: 541).

Yang and Allan's (2007) Service Oriented Architecture for VREs contain some valuable components, which have been included in this researcher's study. The idea of having core pluggable components has been incorporated in the core interface layer (see Figure 3.12a later in this chapter). The ability to expand the VRE through plugging in new services and external services is also a valuable contribution. Yang and Allan's (2007: 541) idea of having a web portal act as presentation layer or interface is also of great value, as well as their idea of using a VLE such as Sakai as interface. Another valuable contribution is their emphasis on standards such as JSR 168, and Web Services Remote Portlets (WSRP) to enable standardised communication between the components in a VRE.

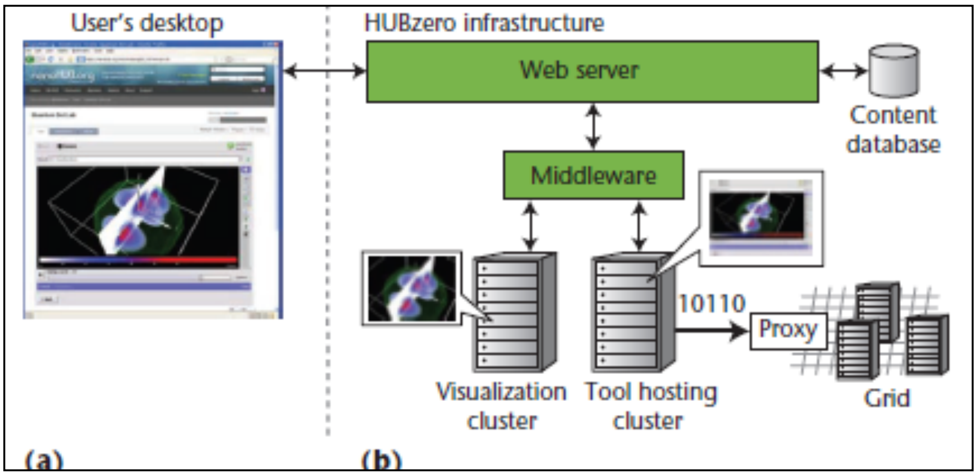
Yang and Allan (2007: 540) also touches on an authentication and authorisation service, which is an essential component that has been incorporated in this researcher's conceptual framework (See Figure 3.12a). The collaboration components in their research support layer, such as chat, e-mail, blogs, and RSS, have also been included in this researcher's possible VRE conceptual framework, but have been divided into two components, namely communication tools (chat, e-mail) and collaboration tools (blogs) (see Figure 3.12a). The information collection/publishing components such as literature search have been included in the searching component (see Figure 3.12a) and the paper publication has been included in the pluggable component "publishing" (see Figure 3.12c).

Yang and Allan (2007), however, does not include anything about the human components, possible hardware components, grid or cloud services, the research cycle, or policy issues in a VRE, which are all important issues this researcher aims to address (see Figure 3.12a-c).

3.5.5 Mclennan And Kennell's (2010) Illustration Of Hubzero

HUBzero, a platform for scientific collaboration, is a cyberinfrastructure that was developed by Purdue University in the USA. HUBzero allows "scientific researchers to work together online to develop simulation and modelling tools" (McLennan and Kennell, 2010: 48). The resulting tools can then be accessed by others through an ordinary web browser and by launching simulation runs on national grid infrastructure, without the need to download or develop any code. McLennan and Kennell (2010: 49) illustrate the HUBzero process as follows (See Figure 3.10):

Figure 3.10: HUBzero's Architecture



(a) Users gain access to interactive, graphical tools by means of an applet in their Web browser on their desktops, and (b), the tools run on a cluster at Purdue University, where requests can be transmitted to more powerful computers in the US national grid infrastructure.

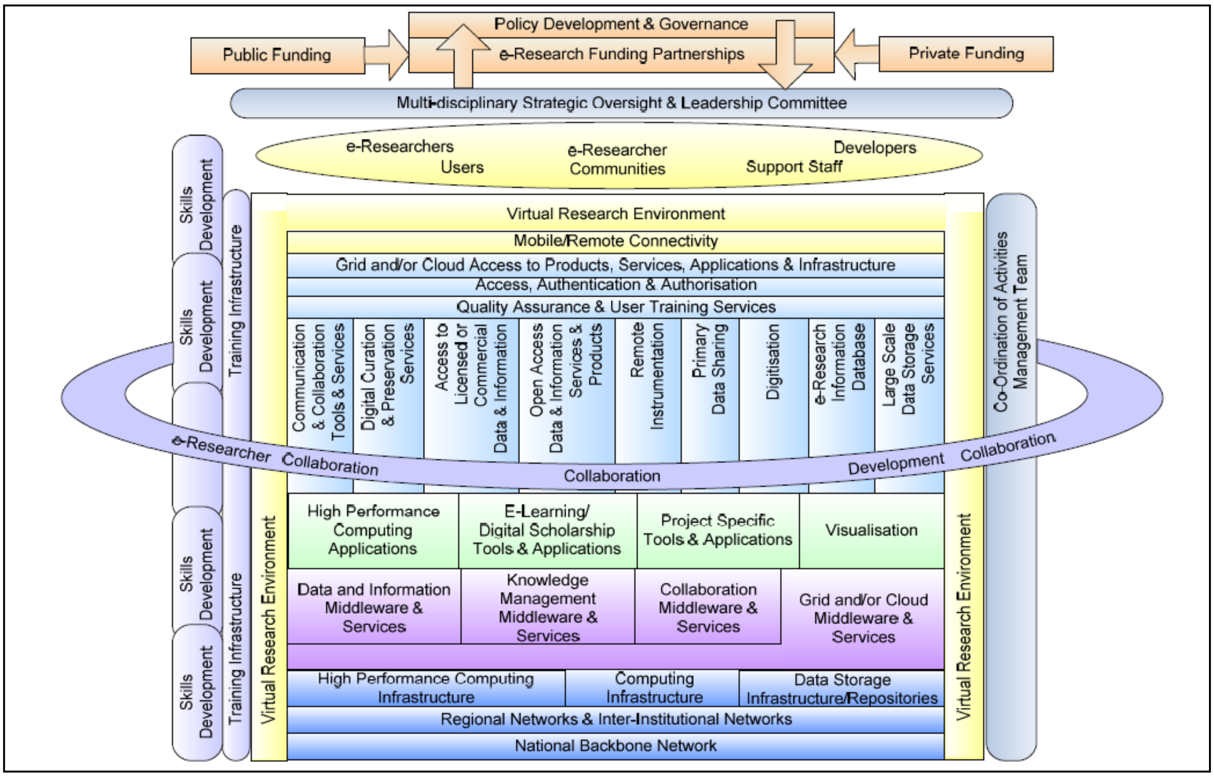
According to McLennan and Kennell (2010: 48), HUBzero supports a growing number of hubs that serve different communities. On the surface, each HUBzero gateway is a website that has been built with 'LAMP' Architecture, comprising "a Linux operating system, an Apache web server, a MySQL database, and PHP web scripting" (Hwang, Dongarra and Fox, 2013: 298). The Rappture toolkit is used to create graphical user interfaces (GUIs) for simulation programmes and specific "middleware for hosting simulation tools and scientific data" (Hwang, Dongarra and Fox, 2013: 298).

Each tool page on a hub has a noticeable button that can be used to start a live session. Clicking on this button opens up an interactive graphic user interface (GUI) for the tool within the user's browser. The simulation tools look similar to Java applets set in the browser, but are actually running on a cluster of execution hosts held "near the Web server and projected to the user's browser using virtual network computing (VNC)" (McLennan and Kennell, 2010: 48-49). Each tool operates in a restricted lightweight virtual environment, which meticulously regulates access to file systems, networking and other server processes. Every user owns an exclusive home directory with standard ownership, access regulations, and quota restrictions (McLennan and Kennell, 2010: 49). The middleware of the HUBzero platform regulates the tool container’s network operations, e.g. it authenticates and directs received VNC viewer and file transfer requests from browsers, to the correct container. At the same time, the middleware keeps an eye on the start time and duration of every connection for book keeping and statistical purposes (McLennan and Kennell, 2010: 49). For example, a tool keeps on running even if it is not being viewed, allowing users to go out of their browsers, and at a later stage go back to their tools by accessing them via their My Hub pages (McLennan and Kennell, 2010: 49). The fact that tools operate in a controlled environment on the hub’s execution cluster and not on the user’s computer, make it possible to authorise them for protected access to high-performance visualisation facilities, as well as running on remote resources. For example, tool containers can be configured to direct jobs through a ‘submit’ server functioning as a secure proxy through which jobs can be directed by the hub to national grid resources in the USA, for instance Open Science Grid, TeraGrid, and Purdue University’s DiaGrid (McLennan and Kennell, 2010: 49).

McLennan and Kennell’s (2010) illustration of HUBzero place an emphasis on simulation and visualisation components, which are valuable components of a VRE framework. Similar to Yang and Allan’s (2007: 541) description of using Java applets to access other VRE components, McLennan and Kennell (2010) make use of portlets to gain access to VRE tools and services. Another valuable contribution of McLennan and Kennell (2010) is the tool hosting cluster, which is kept separately and from which people can request a tool – similar to this researcher’s pluggable components layer (See Figure 3.12c). McLennan and Kennell’s (2010) model also gives access to National Grid Services. Authentication is another valuable component of a VRE, which MacLennan and Kennell (2010) emphasize. Shortcomings of MacLennan and Kennell’s (2010) model are the absence of human components, or hardware components of a VRE. Moreover, no mention is made about the research cycle, standards, protocols, specifications or any policy issues in a VRE model.

3.5.6 Fernihough's (2011: 101) E-Research Implementation Framework For South African Organisations

Figure 3.11: An e-Research implementation framework for South African organisations


Policy Development & Governance
Public Funding
e-Research Funding Partnerships
Private Funding
Multi-disciplinary Strategic Oversight & Leadership Committee
e-Researchers
Users
e-Researcher Communities
Developers
Support Staff
Skills Development
Skills Development
Skills Development
Skills Development
Training Infrastructure
Training Infrastructure
Virtual Research Environment
Mobile/Remote Connectivity
Grid and/or Cloud Access to Products, Services, Applications & Infrastructure
Access, Authentication & Authorisation
Quality Assurance & User Training Services
Communication & Collaboration Tools & Services
Digital Curation & Preservation Services
Access to Licensed or Commercial Data & Information
Open Access Data & Information Services & Products
Remote Instrumentation
Primary Data Sharing
Digitisation
e-Research Information Database
Large Scale Data Storage Services
Co-Ordination of Activities Management Team
e-Researcher Collaboration
Collaboration
Development Collaboration
High Performance Computing Applications
E-Learning/ Digital Scholarship Tools & Applications
Project Specific Tools & Applications
Visualisation
Data and Information Middleware & Services
Knowledge Management Middleware & Services
Collaboration Middleware & Services
Grid and/or Cloud Middleware & Services
High Performance Computing Infrastructure
Computing Infrastructure
Data Storage Infrastructure/Repositories
Regional Networks & Inter-Institutional Networks
National Backbone Network
Virtual Research Environment
Virtual Research Environment

Fernihough (2011: 101-109) compiled a very comprehensive conceptual framework on the e-Research process as a whole, which included VREs, but because of its comprehensiveness, not all the components of this framework will necessarily form part of this researcher's conceptual framework of a VRE.

Fernihough's (2011: 101-109) framework, as indicated in Figure 3.11, consists of the following:

- **An infrastructure or cyberinfrastructure layer (at the bottom in blue):** This layer comprises the physical infrastructure needed to construct an enabling environment for e-Research, forming the foundation for the construction of other layers "to enhance the use of the infrastructure" (Fernihough, 2011: 102). This infrastructure can be divided into the following:

- **National backbone network**:

  "A high speed, large bandwidth network" where all regional and/or inter-institutional networks can link up to. This network also provides the necessary connectivity to other education and research networks and grids internationally (Fernihough, 2011: 102).
- **Regional and/or inter-institutional networks**:

  Networks connecting all institutions in a region together, e.g. all universities in a region, or all research institutions, and then linking to the national backbone network (Fernihough, 2011: 102).
- **HPC infrastructure**:

  Infrastructure that enable "researchers to process large volumes of data at high speeds or do complex analysis" (Fernihough, 2011: 102).
- **Computing infrastructure**:

  Standard computing infrastructure consisting of desktop computers and mobile devices, such as cell phones and tablets, to provide support to researchers so that they can do their research successfully (Fernihough, 2011: 102).
- **Data storage infrastructure / repositories**:

  This infrastructure is used for storage of digital content and assets for future searching and retrieval. Data storage can be done on a national and institutional level (Fernihough, 2011: 102).

- **A middleware and services layer (in purple):**

  This is the communications layer that makes it possible for "applications to interact across hardware and network environments" (Fernihough, 2011: 103). The middleware layer, according to Fernihough (2011: 103), contains the following components:

  - **Grid middleware and services:**

    These afford necessary "access, communication, accounting, security, trust, and co-ordination services" connecting "the (computational and data) resources of the grid and the higher-level services" that utilise them (DSTC, 2004: 3; Fernihough, 2011: 103).
  - **Data and information middleware and services:**

    These contribute services and tools that make the following actions with sizable "heterogeneous distributed data repositories and digital archives" possible:

"indexing, archival discovery, analysis, integration, and management and preservation" (DSTC, 2004: 3; Fernihough, 2011: 103).

  - **Knowledge management middleware and services:**
    These middleware and services utilise the data and information that were generated, archived, indexed etc., to activate knowledge (Fernihough, 2011: 103).
  - **Collaboration middleware and services:**
    These types of middleware offer services and tools to encourage informal and formal, real-time and offline collaborative endeavours between researchers that are located faraway, research communities, as well as resources (DSTC, 2004: 103; Fernihough, 2011: 103).

- **An applications layer (in green):**
  This layer consists of particular applications needed to use the infrastructure and middleware that lies beneath. These can be used as needed. Fernihough (2011: 103) lists the following applications:
  - HPC applications;
  - e-Learning/digital scholarship tools and applications;
  - Visualisation applications; and
  - Project specific tools and applications.

- **Products and services layer (in turquoise)**
  This layer, according to Fernihough (2011: 104-106), comprise those components that researchers may require, but not all institutions. These are:
  - Communication and collaboration components;
  - Digital curation and preservation;
  - Access to licensed or commercial data and information;
  - Open access data, and information services and products;
  - Remote instrumentation (referring to those services that allow the researcher to remotely control instrumentation and equipment);
  - Primary data sharing;
  - Digitisation;
  - e-Research information database (referring to a database of information specifically related to the development of the various components of e-research);

  - Large-scale data storage services;
  - Quality assurance and user training services;
  - Access, authentication and authorisation;
  - Grid and/or cloud access.

- **Users, access and mobile / remote connectivity layer (in yellow)**
  This layer consolidates the applications and policies particularly related to the users and their ability to make use of / access the infrastructure below this layer.
  - **Mobile/remote connectivity** alludes to the service, products and tools that make it possible to access the products, infrastructure or services by mobile or remotely, e.g. cell phones, tablets, as well as the applications needed to run on these devices (Fernihough, 2011:107).
  - **The VRE** forms a barrier layer around the infrastructure (security as well as framework) and relates to the interoperation of an array of online tools, systems and processes across institutional borders with the aim of facilitating or augmenting the research process. The VRE provides researchers with necessary tools and services as well as collaboration facilities for efficient and effective research (Fernihough, 2011: 107).
  - **e-Researchers, e-researcher communities, users, developers, support staff** apply to all persons who can utilise, take part, develop and support the e-Research Framework (Fernihough, 2011: 107).

- **Skills development and training infrastructure layer (in light grey on left side of framework)** concerns the development of the skills process as well as the training infrastructure needed to raise the skills level of researchers, support personnel, and IT specialists, etc.
  - Skills development: A deeper and more rapid adoption of e-Research would necessitate the development of specific skills for different groups of people (Fernihough, 2011: 107).
  - Training infrastructure: The infrastructure needed in order to develop skills. Cloud computing can be utilised to assist with this so that resources can be made available in a dynamically and virtual manner for training, and then re-assigned on conclusion of the training (Fernihough, 2011: 108).

- **Multi-disciplinary strategic oversight and leadership committee layer (in light blue)** is a committee consisting of members from various institutions and multi-disciplines that have a strategic vision of e-Research. The mandate of the committee is to “provide the strategic direction, drive, engagement and co-ordination efforts of research groups involved in e-Research” (Fernihough, 2011: 108).

- **Co-ordination of activities management team (in light blue on right side of framework)**: This is a team tasked with making sure that activities across the various layers of the framework are co-ordinated and that collaboration takes place on all levels (Fernihough, 2011: 108).

- **e-Research funding partnerships (in orange)** concerns the partnership between government, industry and institutions to fund e-Research activities, encompassing the development and implementation of e-Research components nationally and institutionally (Fernihough, 2011: 109).

- **Policy development and governance (in orange)** emphasise the requirement for policies to be reassessed on all levels and particularly “in terms of funding, to ensure co-investment and collaboration” (Fernihough, 2011: 109).

- **Collaboration (light purple band)** was found to be a crucial and necessary driver for e-Research. Collaboration would be expected between stakeholders on all fronts. This also comprise inclusion of collaborative funding, collaborative development, and collaboration in conducting research (Fernihough, 2011: 109).

Many of the components addressed by Fernihough (2011) in her e-Research framework can be included in a conceptual framework for a VRE, but there are components she mentioned, which would not necessarily form part of a generic conceptual framework. Nonetheless, Fernihough (2011: 101-109) identified some valuable components in her e-Research framework, which could also be potentially valuable in a VRE. She saw VREs as overarching the following:

- A cyberinfrastructure layer (consisting of the national backbone network, regional and/or inter-institutional networks, HPC infrastructure, computing infrastructure, and data storage infrastructure/repositories);

- A middleware and services layer (containing grid middleware and services, data and information middleware and services, knowledge management middleware and services, and collaboration middleware and services);
- An applications layer (consisting of HPC applications, e-learning/digital scholarship tools and applications, visualisation applications, project specific tools and applications), which can be used as needed; and
- A products and services layer that researchers may require (consisting of communication and collaboration components, digital curation and preservation, access to licensed or commercial data and information, open access data and information services and products, remote instrumentation, primary data sharing, digitisation, an e-Research Information Database, large scale data storage services, quality assurance and user training services, access, authentication and authorisation, and grid and/or cloud access); as well as mobile and remote connectivity.

The viewpoint of this researcher is that a national backbone network and regional and/or inter-institutional networks form the foundation on which VRE's can be built and operate, constituting the cyberinfrastructure layer. HPC infrastructure and data storage infrastructure/repositories, however, does not necessarily form the foundation, but could be used as required. Fernihough (2011:101) also sees e-researchers, e-researcher communities, users, developers, and support staff as part of the e-Research framework, but places them outside the VRE. This researcher though, regards them as part of the human components layer of a VRE. Furthermore, the idea of a Multi-Disciplinary Strategic Oversight & Leadership Committee would not necessarily be an essential component within a conceptual framework of a VRE, and the 'co-ordination of activities management team' function could be fulfilled by a VRE facilitator (e.g. a VRE Manager and / or a VRE Champion) as an important human component of a VRE.

Fernihough (2011) also does not make a distinction between a core interface layer with essential components and a more flexible pluggable components layer. This researcher is of the opinion that some of the components in the applications layer and products and services layer can be combined and placed in a core interface layer (e.g. communication and collaboration components; access, authentication and authorisation), and some in a layer consisting of RDM components (e.g. digital curation and preservation, open access data services and products; primary data sharing, data storage services (repositories), and a layer consisting of pluggable components (e.g. access to licensed or commercial data and information; open access information services and products; remote instrumentation; digitisation; e-learning tools and applications; visualisation applications; and project specific tools and applications). The grid and / or cloud access could be combined with HPC applications in its own layer, which could be accessed as needed. Furthermore, Fernihough (2011) does not link her model specifically to the research cycle, and consequently misses out on important components such as publishing and peer reviewing.

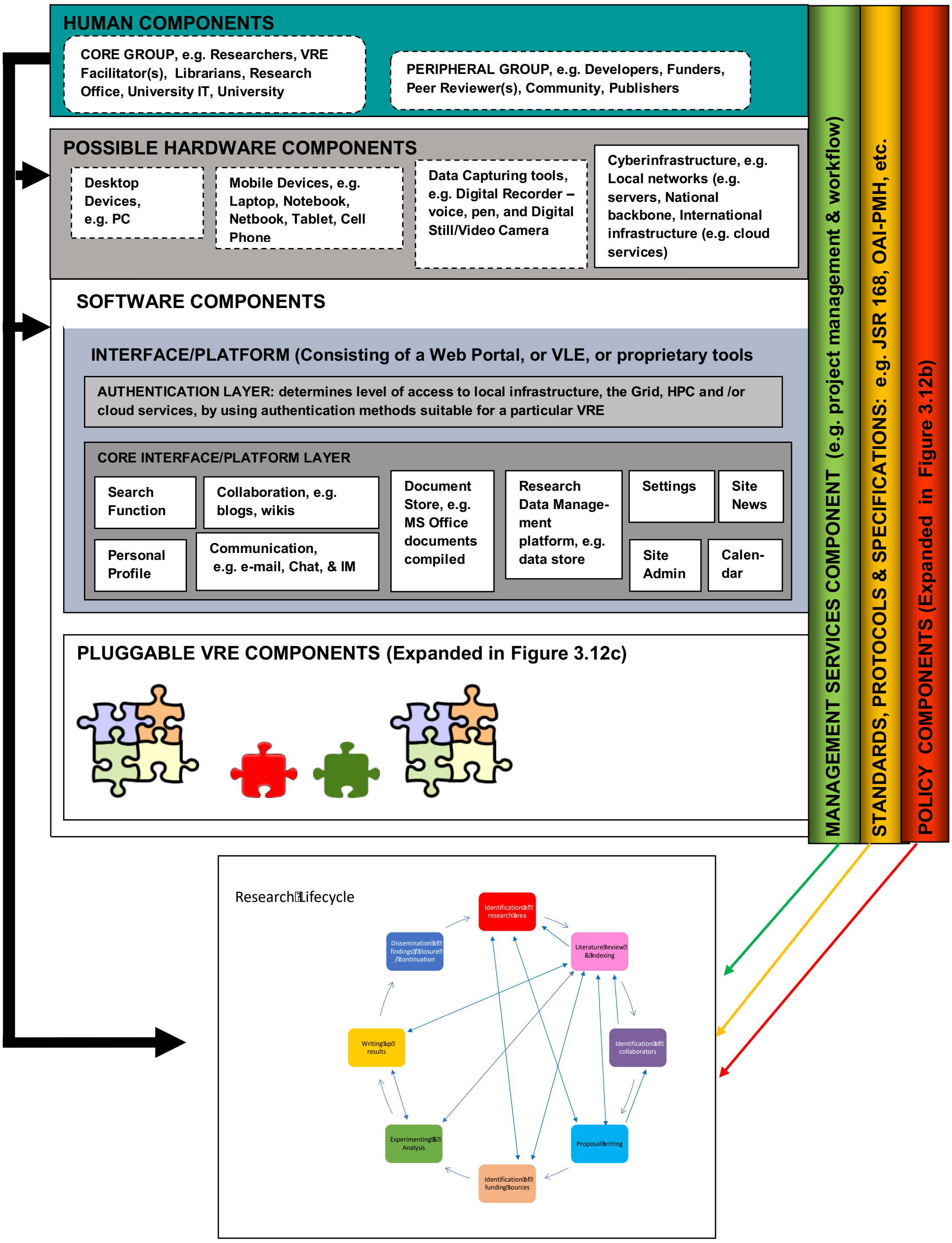
Keraminiyage, Amaratunga and Haigh's (2009b: 129-142) visualised structure of a VRE, De Roure et al.'s (2009) illustration of the myExperiment architecture, Simeoni et al.'s (2008) illustration of the gCube Architecture/framework, Yang and Allan's (2007) Service-Orientated Architecture (SOA) for VRE systems, McLennan and Kennell's (2010) illustration of HUBzero, and Fernihough's (2011: 101) e-Research Implementation Framework for South African organisations, all fall short of providing a possible complete conceptual framework for a VRE. Each of these models, illustrations or frameworks however contribute valuable components that could be included in a possible comprehensive conceptual framework.

The researcher decided to address the shortage in literature by designing a possible conceptual VRE model by combining some of the valuable components provided in the different models, frameworks and illustrations, with components found in other literature, to get a clearer understanding of how these components interact in a research cycle. Also, as discussed in 2.5.3.3., a service-oriented (SO) approach lends itself very well to VREs because of its flexibility to changing user needs, as well as the possibility to expand core services by plugging in new services, or making use of external services, as needed. For the purpose of constructing a conceptual framework of a VRE and its components, a SO-approach has therefore been followed by this researcher.

3.5.7 Proposed Conceptual Model Of A VRE And Its Components

The possible conceptual model of a VRE and its tools as illustrated in Figures 3.12(a-c) consist of a human layer with human components, a hardware layer with possible hardware components, and a software layer comprising software components. These three layers support and impact the research process as it develops through the research cycle.

Figure 3.12a: Proposed Conceptual Model Of A VRE And Its Components

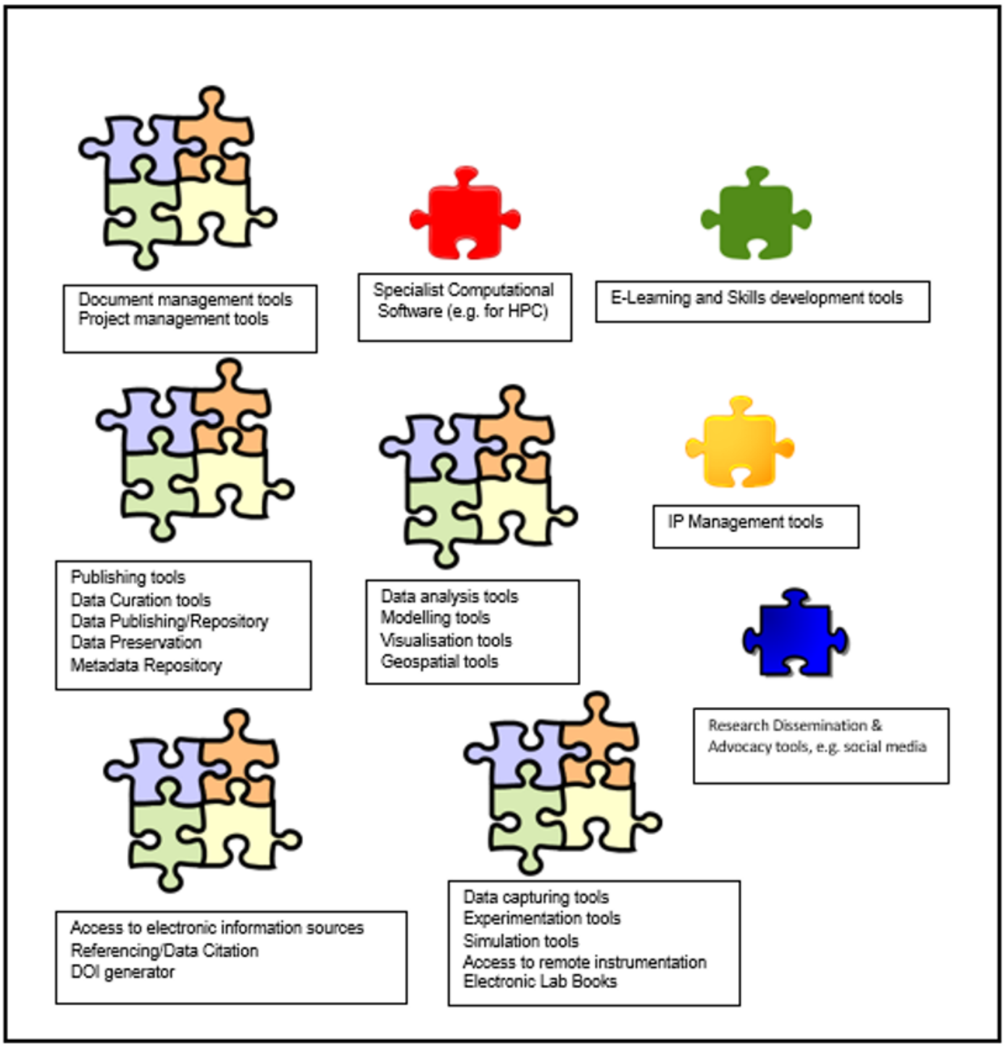


HUMAN COMPONENTS
CORE GROUP, e.g. Researchers, VRE Facilitator(s), Librarians, Research Office, University IT, University
PERIPHERAL GROUP, e.g. Developers, Funders, Peer Reviewer(s), Community, Publishers
POSSIBLE HARDWARE COMPONENTS
Desktop Devices, e.g. PC
Mobile Devices, e.g. Laptop, Notebook, Netbook, Tablet, Cell Phone
Data Capturing tools, e.g. Digital Recorder – voice, pen, and Digital Still/Video Camera
Cyberinfrastructure, e.g. Local networks (e.g. servers, National backbone, International infrastructure (e.g. cloud services)
SOFTWARE COMPONENTS
INTERFACE/PLATFORM (Consisting of a Web Portal, or VLE, or proprietary tools
AUTHENTICATION LAYER: determines level of access to local infrastructure, the Grid, HPC and /or cloud services, by using authentication methods suitable for a particular VRE
CORE INTERFACE/PLATFORM LAYER
Search Function
Collaboration, e.g. blogs, wikis
Document Store, e.g. MS Office documents compiled
Research Data Manage-ment platform, e.g. data store
Settings
Site News
Personal Profile
Communication, e.g. e-mail, Chat, & IM
Site Admin
Calen-dar
PLUGGABLE VRE COMPONENTS (Expanded in Figure 3.12c)
MANAGEMENT SERVICES COMPONENT (e.g. project management & workflow)
STANDARDS, PROTOCOLS & SPECIFICATIONS: e.g. JSR 168, OAI-PMH, etc.
POLICY COMPONENTS (Expanded in Figure 3.12b)
Research Lifecycle
Identification of research area
Literature review & indexing
Identification of collaborators
Proposal writing
Identification of funding sources
Experimenting & Analysis
Writing up results
Dissemination of findings / closure / continuation

Figure 3.12b: Policy Components

- Clear ground rules, e.g. Determine who act as facilitator; Determine the roles in the VRE
- Trust relationships
- Clearly defined objectives
- Mutually agreed project plan/collaborative agreement
- Encouragement of shared interest and enthusiasm
- Intellectual Property (IP) issues across country borders should be dealt with beforehand
- Protection of rights
- Ethical issues must be considered and taken care of
- Proper matching of skills levels and research interests
- Decision on type of interface, type of grid service, and/ or cloud service, pluggable components, standards and protocols
- Negotiations / Decisions on shared access to publications, conference papers (licensing issues)
- Negotiations / Decisions on shared access to research equipment, instruments, and technology
- Negotiations / Decisions on shared opportunities for publishing and presentations
- Regular progress monitoring

Figure 3.12c: VRE Components that can be used or plugged into the VRE



Document management tools
Project management tools
Specialist Computational Software (e.g. for HPC)
E-Learning and Skills development tools
Publishing tools
Data Curation tools
Data Publishing/Repository
Data Preservation
Metadata Repository
Data analysis tools
Modelling tools
Visualisation tools
Geospatial tools
IP Management tools
Research Dissemination & Advocacy tools, e.g. social media
Access to electronic information sources
Referencing/Data Citation
DOI generator
Data capturing tools
Experimentation tools
Simulation tools
Access to remote instrumentation
Electronic Lab Books

3.5.7.1 The Human Components Layer

This layer consists of the various human actors that might possibly want access to such a VRE. Each VRE will have a core group and a peripheral group.

The core group will consist of the following actors:

- **Researchers**, who will use the VRE to support their research process as it develops through the different stages of the research cycle from identification of research area to dissemination/outputs of their research.

- A **VRE facilitator (VRE Manager and/or VRE Champion)** who will play a pivotal role to keep everything and everyone in a VRE together. His/her role will comprise the controlling of permissions (authentication), training users, trouble shooting, evaluating and testing the platform or interface as well as the tool features, identifying resources that can be integrated with the platform, scouting for content that will be relevant, pushing only relevant content to the VRE community, identifying opportunities and making linkages, facilitating the development of relationships and trust, and liaising with the VRE developer(s)/designer(s) (Bowers and Van Deventer, 2012).

- **Librarians,** who can, because of their unique skills, play a very valuable and vital role in ensuring the successful development and optimal use of VREs. Wusteman (2009: 68) identifies three aspects of VREs in which librarians could make a valuable contribution: VRE development, training and use. With regards to VRE development, librarians are well acquainted with the issues around VLEs and have an insight into many of the central issues that will be involved in VRE development, including information access and curation as well as discipline-based knowledge. Some authors such as Candela, Castelli and Pagano (2009: 249) suggest that librarians could also be designers of VREs, although they have an uncertainty whether the entire VRE design, creation and maintenance process, in other words the facilitation of a VRE, could be handled by a single person. Bowers and Van Deventer (2012) question the role of the librarian being the facilitator of a VRE. They see the role of librarians more as populating the VRE with content as well as structuring access to the content. According to Wusteman (2009: 69) though, librarians can play a role in the development/design of a VRE. Librarians, by collaborating with research communities, can ascertain the "user requirements and facilitate user evaluation," which is made possible through trusting relationships and liaisons they have with researchers. The user requirements are then shared with the technologists developing the VRE. Ideas from technologists are also shared with the researchers, which makes the role of the librarian that of a go-between. Candela, Castelli and Pagano (2009: 248), in turn, stress the role of a librarian as populating the VREs with information sources, "aggregating and rearranging knowledge in different subject areas," as well as creating tools and interfaces that will allow for the searching and usage of the information resources.

Due to the complexity of the VRE environment, the need for e-Research literacy is expanding, and librarians can play a constructive role in training researchers to use and manage VREs, as well as the tools within them (Wusteman, 2009: 69). Librarians can also play a valuable role in ensuring that appropriate information-related standards and solutions are used in VREs, e.g. with regards to the usage of metadata (Wusteman, 2009: 69). Librarians could furthermore perform an important role in checking "that open access publications do not violate any third-party rights before publication" and in advising researchers on copyrights and licensing issues (Carusi and Reimer, 2010: 54, 73). Another role that librarians could play is that of collecting, curating, preserving, maintaining and archiving of various digital assets such as software repositories, the research workflows, research data, and research outputs (publications) (Candela, Castelli and Pagano, 2009: 249).

- **The University Research Office** acts as intermediary between funders and researchers, and provide the necessary information with regards to information needed for application for funding (e.g. RMD plans).

- **The University IT Department** should take responsibility for the roll-out and maintenance of the necessary IT infrastructure.

- **The University Executive** could play a pivotal role in providing and applying policy and strategy with regards to RDM at an institution, but also with regards to VREs in general. The Executive would also be responsible for the provision of necessary resources (e.g. funding and staff) for the management of research data.

The peripheral group will consist of the following human actors:

- The **developers / designers** of the VRE, who will need access to all levels of the VRE to develop, build and sustain its features.

- The **funders** of research projects that use the VRE will also need access to certain parts of the VRE to keep track of the progress of the project(s) they are funding.

- **The peer reviewers.** Peer reviewing form an essential part of the research process, and researchers taking part in a specific VRE project need to have access to each other's work in order to share and comment on one another's work. They also need to be able to communicate via the different channels of the VRE. Peer reviewers ensure that the data in the VRE are of high quality. Reviewers from outside the VRE project should be able to have access to certain areas of the VRE, e.g. the repositories, but should only have reading rights. They must nonetheless be able to communicate via the communication channels in the VRE.

- **The community.** Members of the community (public) might sometimes also have access to certain parts of the VRE, which can be repositories, wikis or blogs.

- **The publishers**. Publishers might need access to the underlying research data of the articles they publish. The availability of the data on which an article is based is increasingly becoming a prerequisite when publishing in a journal. A VRE can provide the necessary access to the data.

3.5.7.2 Hardware Components Layer

This layer consists of the various hardware components that can potentially be chosen by the human components in a VRE configuration, or to access a VRE. These components can be grouped in four categories: desktop services, e.g. personal computers (PCs); mobile devices, e.g. laptop computers, notebook computers, netbooks, computer tablets, or cell phones; data capture and output devices, e.g. digital still cameras, digital video cameras, and digital recorders such as digital pens and voice recorders; as well as cyberinfrastructure, including local networks (e.g. servers), the national backbone, and international infrastructure (e.g. cloud services as tools to assist in accommodating the vast amounts of data that will need to be managed). It should be noted that this is not a definitive list and more hardware components can be added as needed.

3.5.7.3 Software Components Layer

The software components layer itself consists of the following sub-layers: an interface or platform layer, and a layer with components (applications and services), which can be plugged into the VRE as needed. The **interface or platform layer** normally forms the front-end of the software component layer of a VRE. This is the part of a VRE that is seen and accessed by the human components. The interface is often in the form of a web portal, or reconfigured VLEs, or shelf-ready tools such as HUBzero, myExperiment, ResearchGate, etc.

As part of the interface or platform, there is also an **authentication layer** to determine the level of access a human component can have to the software layer. The various human components will have different access rights that are determined through a registration process and logins and passwords. This layer is absolutely essential as a security mechanism for the VRE. Allan (2009: 115-121) lists a number of potential primary authentication methods:

- Trust authentication, “where the system assumes that anyone that can connect to it” is authorised to access it with any user name they specify, in other words anonymous access (Allan, 2009: 116);
- Password authentication, where the user is given a special username and password to access the VRE, for example through a portal interface (Allan, 2009: 116);
- LDAP authentication, which is used to validate existing user name/password pairs;
- PAM authentication, which is similar to the password authentication, but different in that it uses pluggable authentication modules (PAM) as authentication mechanisms, and is normally applied to validate user name/password pairs;

- Ident-based authentication where the client's user name is obtained from their desktop operating system and then uses a map file that lists the permitted user names on the VRE system to determine access;
- Simple Ident authentication over TCP/IP can be found in Unix-like operating systems. Most of these systems come with an Ident server that listens on TCP and UDP port 113 by default. In an VRE environment, the VRE systems can "interrogate the server on the host of the connecting client and theoretically determine the username on the user's desktop operating system for any given connection," but this is very dependent on the integrity of the client (Allan, 2009: 117);
- Ident-map authentication can be done after the username on the client operating system that initiated the connection, was established. The VRE application verifies if a client is allowed to connect with that particular identity using a map file;
- Kerberos authentication is used for distributed computing over a public network;
- GSSAPI authentication "is a protocol for secure authentication," which is defined in industry standard RFC 2743. It provides automatic authentication (single sign-on) for systems that support it. PostgreSQL, a popular open source database used in some VREs, supports GSSAPI (Allan, 2009: 118);
- GRID Security Infrastructure (GSI) according to Allan (2009: 118-119) is a component of Globus middleware. Clients, services and resources are all "identified by certificates, which are issued to them by a trusted entity" called a certification authority, through a formally-defined and legally-binding authentication process. This is necessary in order for them to be accepted by grid resources. GSI offers a delegation capability that is valuable in VREs. In a VRE that accesses quite a lot of resources where each requires mutual authentication, or where there is a need to have agents (distant or local) requesting services on behalf of the client, the necessity to re-enter the client's password can be circumvented by constructing a proxy (consisting of a new certificate with a public key and private key containing the owner's identity);
- MyProxy authentication is useful for various services in a VRE that need to be invoked on the client's behalf, and is described as "a proxy certificate repository used to enable pervasive access to resources from web portals" (Allan, 2009: 120);

- Shibboleth is described by Allan (2009: 120) as an architecture that enables organisations to construct single sign-on environments that would permit clients to access web-based resources by means of a single login.

The core interface / software layer consists of fixed components that are part of the standard configuration of the specific tool used and could vary, but are normally things such as a search function; a personal profile; collaborative writing tools such as blogs and wikis; communication tools such as instant messaging, chat, and e-mail; a document store where MS Office documents can be compiled and stored; a RDM component, e.g. a research data store (more components could be added to enhance the RDM functionality); a settings function; a site news function; a site admin function; and a calendar. Sometimes it might include some of the components that have been listed in the pluggable components layer. The RDM component and its importance is expanded upon in Chapter 5.

The bottom layer of the software components comprises various software components (services and applications) (See expansion in Figure 3.12c), which can be used or plugged into the interface/platform component determined by the needs of each VRE community/project. These components can vary, but an overview of the literature on VREs shows that the following components might be needed (grouped together by related function):

- Document management tools, and project management tools;
- Specialist computational software (for example to use for HPC, and sequencing);
- E-learning and skills development tools;
- Publishing tools, data curation tools (e.g. data management planning tools), data publishing tools (e.g. data repositories), data preservation tools, metadata store;
- Data analysis tools, visualisation tools, modelling tools and geospatial tools;
- Intellectual property management tools;
- Access to electronic information sources, and referencing, and Digital Object Identifier (DOI) generator;
- Experimentation tools, simulation tools, access to remote instrumentation and electronic lab books (Chambers, 2002: 389-390; Di Muro and Saunders, 2008, Interview with Van Till and Dovey, JISC on 1 June 2010 at HEFC Building, London; Keraminiyage, Amaratunga and Haigh, 2009(b): 133-134; Klyne, 2006;

Myhill, Shoebridge and Snook, 2009: 230; Pienaar and Van Deventer, 2009; Scratchpads: biodiversity online, n.d.; Sergeant, Andrews and Farquhar, 2006; Van Deventer et al., 2011; and Voss and Procter, 2009: 175, 179).

3.5.7.4 Management Services Component (Vertical Layer In Green)

The management services component confers automatic behaviour to the whole VRE across the different layers and components by utilising standards, protocols and specifications in service invocation. This component is also used to monitor and execute the individual stages/workflow across the entire VRE. According to Simeoni et al. (2009), this is done through the initiation and dynamic configuration of workflows that oversee service invocations and runtime specifications. Deployment strategies are controlled through brokering and matchmaking algorithms.

3.5.7.5 Standards, Protocols And Specifications (Vertical Layer In Amber)

The various sub-layers within the software components layer are held together by interoperable standards, protocols and specifications, e.g. JSR 168, OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), Z39.50, and SRU/SRW, API, etc. These standards, protocols and specifications help the various software components to communicate with each other and to exchange data with each other. A **protocol** can be defined as “a set of rules or conventions formulated to control the exchange of data between two entities desiring a connection” (Kumar, 2009). Basic ingredients of a protocol, according to Kumar (2009), comprise “data format and signal levels, control information coordination and error handling, [as well as] timing.”

Standards can be defined as “a prescribed set of rules, conditions or requirements concerning definitions of terms; classification of components; specification of materials, performance or operations; delineation of procedures; or measurement of quantity and quality in describing materials, products, systems, services or practices” (National Standards Policy Advisory Committee, 1978: 6). Various types of standards are listed by Allan (2009: 108): Java standards for programming language technology, classes and standard patterns, e.g. JSR 168 (portlet-1), JSR 286 (portlet-2) and JSR 170 (repository); browser-based web technology standards, e.g. AJAX, CGI, JSP, JavaScript and Portlets; web services standards e.g. SOAP, WSDL, WSRP, UDDI, XML and pub-sub pattern; security standards, e.g. TLS, SSL, Kerberos, GSI, SAML and X.509; metadata standards e.g. MARC and Dublin Core; database management standards, e.g. SQL, JDBC and Hiberbate; data discovery access standards, e.g. Z39.50, OAI-PMH, SRW/SRU, OpenURL and OpenSearch; workflow standards, e.g. SCUFL and BPEL.

An example of a **specification** is API (Application Programming Interface), which can be defined as an “interface (consisting of pieces of programming code) implemented by an application which allows other applications to communicate with it” (Kashyap, 2010).

3.5.7.6 Policy Components (Vertical Layer In Red)

Every VRE has a list of important policy components that have to be considered to ensure the successful operation of the VRE. These policy issues are closely related to the human components layer and will have a profound impact on that layer, but also on the functioning of the other layers and the choice of components used. The policy components are expanded in Figure 3.12b; they are: clear ground rules, e.g. who act as facilitator; determining the roles in the VRE; trust relationships; clearly defined objectives; a mutually agreed project plan/collaborative agreement; the handling of intellectual property issues across country borders beforehand; protection of rights and indigenous knowledge rights; the consideration and handling of ethical issues; proper matching of skills levels and research interests; decision on type of interface, type of grid service and/or cloud service, pluggable components, standards and protocols; negotiations/decisions on shared access to publications, conference papers (licensing issues); negotiations/decisions on shared access to research equipment, instruments, and technology; negotiations/decisions on shared opportunities for publishing and presentations; and regular progress monitoring.

3.5.7.7 Research Cycle

At the bottom is the research cycle and each of its components that the VRE with its different components (human, hardware, software, standards, protocols and specifications, management services and policy) will aim to support and enhance. The research cycle chosen for this study consist of the researcher's adapted version of Pienaar and Van Deventer (2009), and Van Deventer et al.'s (2009) research cycle, consisting of the following components that function iteratively and not necessarily in a cycle: identification of research area; literature review and indexing; identification of collaborators; proposal writing; identification of funding sources; experimentation and analysis; writing up results; and dissemination/output of findings.

3.6 SUMMARY

The first part of this chapter touched on the development of VREs around the world, focusing specifically on the UK, the USA, the Netherlands and Germany.

An overview of the UK showed that VREs funded by JISC developed in three phases, each with their own strands, while the British Library and the Technical Computing Group at Microsoft established a joint venture to develop the Research Information Centre (RIC).

An overview of the USA showed that their VRE programmes developed very ad hoc, and also used different terminology, e.g. collaboratories and science gateways. Science gateways seem to have been the main driver of VRE programmes in the USA. Other VRE developments in the USA comprise HUBzero on the one hand, and Alzforum, Schizophrenia Research Forum, PD (Parkinson's disease) Online and StemBook, on the other hand.

In the Netherlands, the SURF organisation has been the driving force behind their VRE programme, with the SURFfoundation as the user-facing division of SURF. SURFshare is a SURFfoundation project which was pioneered in 2007. It had six major themes, of which collaboratories was one. SURFnet is another division of SURF responsible for development of the technology. The literature showed that both in 2007 and in 2008, calls for tenders for collaboratories were issued by SURFshare, with tender projects in 2007 focusing on the development and implementation of short-term projects, followed by the widening and deepening of the experiences of 2007 in the 2008 tender projects, for example by increasing experience in more disciplines/institutes, as well as gaining more insight in user experiences and about the impact of the collaboratories on their work. The collaboratory projects funded by the SURFshare programme were: Collaboratory for Evidence Based Critical Reviews, Hublab, Tales of the Revolt, Testweeklab, and Virtual Knowledge Studio (VKS). Programmes not funded by SURFshare included Alfalab, Collaboratory.nl, Digital Collaboratory for Cultural Dendrochronology in The Low Countries (DCCD), eLaborate, LabsOnline, and PARTNER.

The German research foundation Deutsche Forschungsgemeinschaft (DFG) funded and drove the majority of Germany's VRE programmes. DFG issued specific VRE calls in 2008 and 2009. Fifteen bids were received in the 2008 call; six of which were subsequently funded in 2009. In the 2009 round, 48 bids were received, and sixteen projects were subsequently approved in 2010.

The researcher then discussed the similarities and differences between the VRE programmes of these four countries by hand of organisational-, technical-, functional-, policy/legal/financial- and cultural aspects. This was followed by an overview of the concept of research cycles as described in literature. After this, a layout was given of the various components that a VRE can consist of, and this was then matched to the stages of a research cycle, together with possible VRE tools (see also Addendum A).

Finally, all the elements of this chapter were brought together in a possible conceptual VRE framework consisting of a human layer with human components, a hardware layer with possible hardware components, and a software layer, comprising software components. The human components layer was shown to consist of researchers, a VRE facilitator, developers, funders, librarians, peer reviewers, as well as the community. The hardware components layer was shown to consist of the various hardware items that can potentially be chosen by the human components in a VRE configuration. The software components layer was shown to consist of sub-layers: an interface or platform layer, and a layer with components (applications and services) that can be plugged into the VRE as needed. The interface or platform layer was further divided into an authentication layer and a core interface/software layer. The various sub-layers within the software components layer were shown to be held together by interoperable standards, protocols and specifications, as well as management services. The successful operation of a VRE was finally shown to depend on a list of important guidelines and policy issues that will have to be considered to ensure the successful operation of any particular VRE.

The literature study showed that it is possible to compile a conceptual framework of a VRE, in theory. These ideas, however, still have to be tested in practice; therefore, the researcher decided to focus on a case study in the empirical part of this study.

The next chapter will focus on the concept of RDM.

CHAPTER 4
RESEARCH DATA MANAGEMENT (RDM)

4.1 INTRODUCTION

Internationally, research data are recognised as vital resources that have value and need to be preserved for future research (Donnelly, 2015; High Level Expert Group on Scientific Data, 2010: 2; Sandland, 2009: 1; Wolski and Richardson, 2011: 1-2). This can be seen through the numerous initiatives in research institutions, such as universities, research centres, and research laboratories, as well as disciplines across the globe (Beitz, Dharmawardena and Searle, 2012: 1-17; Bradley, 2013: 26; Treloar, Choudhury and Michener, 2012: 174; Norman and Stanton, 2014: 253-262). The value of RDM is also increasingly being realised by governments, funders and publishers (High Level Expert Group on Scientific Data, 2010: 2; Sandland, 2009: 1; Treloar, 2009: 127; Wolski and Richardson, 2011: 1-2). This places a huge responsibility on higher education institutions to ensure that their research data are managed in such a manner that they are protected from substantial reputational, financial and legal risks in the future.

The researcher deemed it necessary to first understand the process of RDM, by looking at what is meant by the concepts data and research data, as well as the various concepts that describe the management of research data, and how each relates to the concept of RDM. The researcher discusses the concepts data curation, data stewardship, data governance, data archiving, and data management. This is followed by a definition of what RDM is. Next, an overview is provided of a number of international developments with regards to RDM, followed by a comparison of the similarities and differences in these different approaches to RDM. The South African situation on RDM is deliberated upon next, with a discussion on government initiatives, national collaborative initiatives, initiatives at higher education institutions, other initiatives and potential partners. Next, the researcher explores the concept of a research data cycle by comparing a number of cycles from literature. Following this, the researcher discusses the different stages of a research cycle as well as the corresponding processes that take place in each, and the potential role that the various stakeholders can play in each. Processes that take place throughout the whole research lifecycle are also discussed. The concept of big data is deliberated upon next, followed by an investigation into the value of RDM. At the end of the chapter, an overview is provided on the developments regarding RDM at the University of Pretoria, South Africa - the location of the case studies this study focuses on.

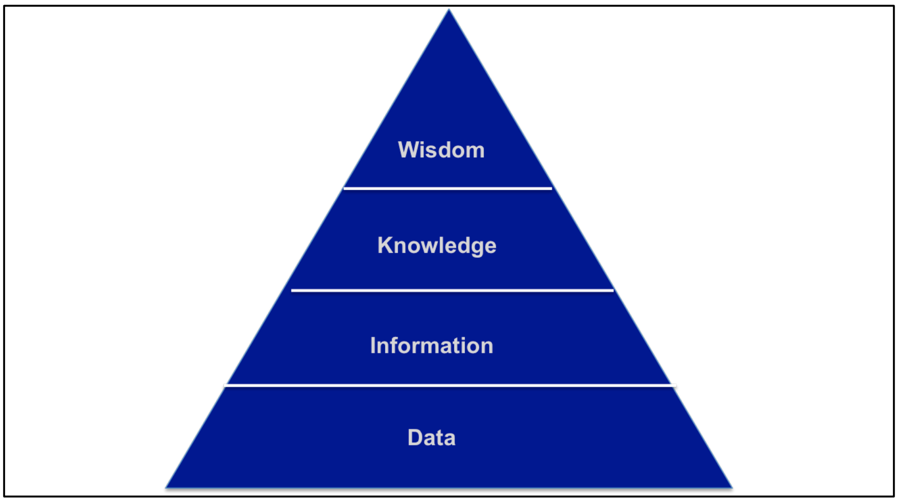
The key concepts data, research data and RDM are addressed next.

4.2 KEY CONCEPTS

4.2.1 Data

The concept ‘data’ can mean different things for different disciplines. The computer science discipline is referred to in Denmark as the “science of data”, and data in this discipline are associated with data processing (Nielsen and Hjørland, 2014: 223). In information science and knowledge management, data are very often discussed in terms of their relationship to the information and knowledge hierarchy. In this, context data are often seen as the raw materials of information processing and knowledge acquisition. This led to the formulation of a hierarchy of data, information, knowledge and wisdom (DIKW). The first version of this hierarchy was proposed by Ackoff (1989: 3-9), and can be seen in Figure 4.1.

Figure 4.1: Hierarchy Of Data, Information, Knowledge And Wisdom (DIKW)



In the DIKW model, it is suggested that there is a hierarchy of types built on the foundation of data. Located at the top is wisdom. Descending from wisdom are knowledge, information and then at the base, data. Each of these types includes the ones that lie below it (Ackoff, 1989: 3). Ackoff (1989: 3) defines data as the "symbols that represent properties of objects, events and their environments. They are products of observation. To observe is to sense." He thereafter refers to the technology of sensing and instrumentation, which is highly developed, to generate data. The acquisition of data can, however, "be generalized well beyond automatic instruments" (Fricke, 2009: 133). "When, for example, a person fills in a form giving their name, address, age, social security number – these inscriptions are data" (Fricke, 2009: 133). This data form the base of the model. Next up in the hierarchy is information, which is described by Ackoff (1989: 3) "as relevant, or usable, or significant, or meaningful, or processed data." Data are processed as an answer to an enquiry and then become information. The difference between data and information is consequently seen by Ackoff (1989: 3), as functional and not structural. Furthermore, according to this view, information can also be inferred from data.

A number of authors have, however, been very critical of the DIKW model. Some of the most vocal critics have been Fricke (2009) and Ma (2012). According to them, the DIKW model presents a theory of knowledge, based on positivism, which has its origin in a tradition of empiricism, while metaphorically depending upon data processing rather than a communicative model for constructing its concept of information (Fricke, 2009: 136-137; Ma, 2012: 721). Ma also notes that the epistemological assumptions in this model have resulted in disregarding the cultural and social features of the constitution of information; in other words, how something is regarded as information or not. It also resulted in "the unquestioned nature of science in research methodologies" (Ma, 2012: 716). Zins (2007: 481), in turn, mentions the much earlier research by Capurro, who already in 1978 expressed a general scepticism regarding the concept of data, when he stated that the idea that "information is set together out of data and [that] knowledge comes out from putting together information," is a fairy tale. His criticism is in line with what later authors Fricke (2009) and Ma (2012) found.

Machlup (1984: 646-647) states that data should be considered as a relative concept. According to him, the word "data" comes from the Latin word 'dare', which means 'to give', and the word 'datum', which means 'the given'. The plural is the word 'data', which means 'the givens.' Data, according to Machlup (1984: 646-647) will mean different things for different people. Data are therefore considered as relative to a process in which something is deemed as 'given' (Nielsen and Hjørland, 2014: 225). Nielsen and Hjørland (2014: 225), in line with Fricke (2009), Ma (2012), Capurro (1978) and Machlup (1984) come to the conclusion "that the nature of data is fundamentally a problem in the philosophy of science." This leaves researchers with no positive definition of 'data' as an alternative to empiricism. The definition of data by Kaase (2001: 3251) is however broad enough to fill this need. Kaase (2001: 3251) suggests that "data is information on properties of units of analysis." In other words, divergent research projects will have different units of analysis, as well as variable levels of relevance for information about disparate properties. These properties, according to Nielsen and Hjørland (2014: 225), are "not just measures, but are any kind of characteristics of an object being subjected to investigation." Included in this definition is the notion that data are always recorded on the premise of specific interests, viewpoints, technologies, and established practices that shape their meaning, application and practical use in dissimilar contexts (Nielsen and Hjørland, 2014: 225).

The definition of Kaase (2001: 3251) forms the basis for the approach to data followed by this thesis. The concept of 'research data' also reflects these differing viewpoints on data.

4.2.2 Research Data

Data in the context of VREs include research data, but also other types of data. A taxonomy of the various types of data found in a VRE was drawn up in Table 4.1 to give a clearer understanding of each of these data types.

Table 4.1: Data Types Found In A VRE (See also 3.2.3, 3.4.2, 3.5.6 and 3.5.7)

Data Type	Sub-Data Type
Research Data (Experimental Data)	Numeric Data Visual Data: still, moving, and animation Textual Data Audio Data
Referencing Data	List of Literature Consulted List of Literature Created List of Datasets
Funding Data	
Collaboration Data	
Administrative Data	

Research data is defined by the National Science Foundation in the USA as “the recorded *factual material* commonly accepted in the scientific community as necessary to validate research findings. This includes original data, but also metadata (e.g. experimental protocols, code written for statistical analysis, etc.)” (NSF, n.d.). The Office of Management and Budget of the White House Administration, in turn, defines research data in their OMB Circular A-110 as: "the recorded *factual material* commonly accepted in the scientific community as necessary to validate research findings" (USA. White House, Office of Management and Budget, 1999). And, according to the Edinburgh University Data Library Research Data Management Handbook (2011), research data, “unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.” These definitions can be synthesized into the following definition: Research data are information on properties of units of analysis, that is collected, observed, created, or generated for the purpose of analysis to validate research findings.

In Table 4.1, it was mentioned that research data could typically be in the form of numeric data, visual data, audio data, and textual data. Numeric data can be described as data that consist of positive and negative numbers, decimal and fractional numbers, as well as whole numbers (integers) (Microsoft TechNet, 2015). Numeric data can further be divided into discrete and continuous numeric data. Discrete data constitute items that can be counted (listed), for example 0, 1, 2, 3 etc. This list of items can be fixed, in other words be finite, or can go on to infinity (be countably infinite). Continuous data constitute measurements, in other words, their potential values cannot be counted and can only be illustrated using intervals (Rumsey, 2015).

Visual data, according to Vannini (2008: 929-930), “include iconic objects and the symbolic meanings that people attach to these,” while visual data analysis focuses on “iconic signs, which can consist of ‘still or moving pictures (e.g., advertisements, videos, film), drawings, paintings, maps, and other images’.” It can also deal with “public behavior (especially nonverbal interaction), material culture, landscape, and the human body and its adornments.” Visual data are described by Grady (2008) as “any visually perceptible object of interest to, or produced by, human beings,” as well as “visually perceptible artefacts that record human doings of one kind or another.” Visual data, in other words, encompass all features “of the physical universe that can be perceived, either directly or indirectly”, for example, a representation of an earthquake by a remote sensing device (Grady, 2008). Visual data can also refer to varied types of images and pictures, “consciously constructed to either record or represent the world” (Grady, 2008). Examples of visual data are photographs, videos, graphic representations in the form of charts, sketches, animation and maps.

Textual data are described by Benoit (2011: 526) as “systematically collected material” comprising “written, printed, or electronically published words, typically either purposefully written or transcribed from speech” (Benoit, 2011: 526).

Audio data are information that consists of the capturing or recordings of sounds. These can be in analogue or digital format. Audio data in analogue form are normally captured by an analogue audio recording device, for example, a conventional tape recorder, which then “captures the entire sound wave form and saves it in analog format on a medium such as magnetic tape” (Murray and Van Ryper, 1996). Audio data in digital form are normally captured through devices such as computers, or mobile devices such as digital recorders or mobile phones, etc. Instead of recording the entire wave-form as analogue devices do, a digital audio recording device “captures a wave form at specific intervals, called the sampling rate” (Murray and Van Ryper, 1996). “Each captured wave-form snapshot is converted to a binary integer value and is then stored on magnetic tape,” disk, or even in the cloud (Murray and Van Ryper, 1996).

Referencing data in this study can be described as information that attributes and acknowledges the original author or source of the data or information. This can consist of a list of literature consulted to generate the data, as well as a list of literature that was generated from the data (e.g. journal articles, reports, books, theses, etc.). Referencing data can also consist of a list of secondary datasets (created by other authors or sources) that were consulted in order to generate or create new datasets.

Funding data contain information about the sources of funding that was used to conduct the research (from creation and generation of data, processing of data, analysis of the data, to the writing up and publishing of the data). Collaboration data comprise personal information and details about the co-researchers on a specific research project. Administrative data in a VRE could include protocol development information, information on the protocol defence, or even registrations for clinical trials, ethical clearance forms, and letter and/or signed permission forms from respondents, etc. Complex data in the context of this study consist of a combination of two or more of these types of data.

This complexity in the concepts ‘data’ and ‘research data’ as discussed, necessitates the need for the proper management of research data. This will be discussed next.

4.2.3 The Concept ‘Research Data Management’

There are a diverse number of competing concepts used in literature to describe the activity of managing research data: data curation, data stewardship, data governance, data archiving, and data management.

4.2.3.1 Data Curation

A review of literature shows that there is no clear-cut definition of the concept data curation. The UCLA Library defines data curation as “the active and ongoing management of data throughout its entire lifecycle of interest and usefulness to scholarship” (UCLA Library, 2014). According to the University of Illinois Graduate School of Library and Information Science, "data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time" (Cragin et al., 2007). The University of Minnesota Libraries regard curation specifically as a value-adding activity. They define data curation as “the value-added activities and features that stewards of digital content engage in to make digital content meaningful or useful” (University of Minnesota Libraries, 2014). The Digital Curation Centre (DCC) in its ‘Charter and Statement of Principles’ use the term ‘digital curation’ for the curation of data, and also mention the aspect of adding value. They define it as “maintaining and adding value to a trusted body of digital research data for current and future use,” but then adds the aspect of “active management of data throughout the research lifecycle,” which corresponds with Cragin et al.’s (2007) definition (DCC, 2014c).

Lord and MacDonald (2003: 12) stresses the management and promotion of “the use of data from its point of creation, to ensure it is fit for contemporary purposes and available for discovery and re-use,” while the University of California San Diego regards “archiving and preservation as subsets of the larger curation process, which is much broader, planned, and interactive.” They define data curation as “managing data to ensure they are fit for contemporary use and available for discovery and re-use” (UC San Diego, 2014). Dempsey (2007), in turn, describes data curation as the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Dempsey (2007) lists the value that data curation activities add: enabling data discovery and retrieval, maintaining its quality, adding value, and providing for re-use over time. Data curation, according to Dempsey (2007), “includes authentication, archiving, management, preservation, retrieval, and representation.”

The Data Conservancy in its data conservancy stack model sees data curation as a process that adds value to the data lifecycle, and then positions data curation within data management as one of the layers of data management (Choudhury, 2013). In other words, data curation is seen as a subset of data management, which adds value to the data management process. This is the approach this study followed.

The above definitions of data curation can be synthesized in the following definition: data curation is the active and ongoing activities that data stewards engage in to add value to research data throughout its entire lifecycle so that the data are meaningful and useful to scholarship, research and education, and available for discovery and re-use.

4.2.3.2 Data Stewardship

Rosenbaum (2010: 1444) and Diamond, Mostashari and Shirky (2009) regard the concept of data stewardship as "rooted in the science and practice of data collection and analysis and reflects the values of fair information practice." Data stewardship according to Rosenbaum (2010: 1444) signify an approach to the management of data, and specifically gathered data that can identify individuals. She describes data stewardship "as a collection of data management methods covering acquisition, storage, aggregation, and de-identification, and procedures for data release and use" (Rosenbaum, 2010: 1444). The US Geological Survey (USGS), in turn, describes stewardship as being equal to taking "responsibility for a set of data for the well-being of the larger organization, and operating in service to, rather than in control of, those around us" (USGS, 2013).

Haines (2012) regards stewardship as a **tactical** function, in other words short-term, specific, and local. Stewardship tasks, according to her, are all executed against specific business terms or data elements. She lists examples of tactical data stewardship tasks as:

- "Defining the data: Identifying key data, gathering definitions, documenting allowable values;
- Defining business rules: For creation of data, for usage of data, for derivation of data;
- Documenting data sources: Identifying system of record/system of recommendation;
- Setting data quality targets: Fit-for-use thresholds;
- Metadata identification/documentation; [and]
- Remediation of data issues."

Data stewardship, in summary, can be described as a specific approach to data management; it is about taking responsibility for data sets, and is a tactical function that is executed against specific data criteria (Haines, 2012).

4.2.3.3 Data Governance

Haines (2012) classify data governance as a *strategic* function. According to her, it is strategic in the sense that it is *long-term*, *general*, and *global*. She lists the following examples of data governance tasks:

- Creation of a structure for participation (consisting of the committees, working groups, and councils for the data governance program);
- Defining the goals and principles (used to guide the data governance programme);
- Establishing a communications plan;
- Defining the policies and processes (used to implement data governance); and
- Define the roles and responsibilities (stipulates the rights and obligations of the participants in data governance).

For Haines (2012), data governance is **not** about the data per sè. "It's about the **people, policies**, and **processes** to manage an asset that *happens to be* data."

Rouse (2007) defines data governance as referring "to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise," while the Dictionary of Data Management (DAMA, 2011) defines data governance as "the exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets" (DAMA, 2011).

In summary, data governance is therefore more concerned with the people managing the data, which includes goals, policies, shared decision making, planning, strategies, and processes followed.

4.2.3.4 Data Archiving

Data archiving in a computation context, according to Müller (2009), "refers to the storage of electronic documents, data sets, multimedia files, and so on, for a defined period of time." The report of the CODATA Workshop on Archiving Scientific & Technical (S&T) DATA, 20-21 May 2002, describes data archiving as "primarily a program of practices and procedures that support the collection, long-term preservation, and low-cost access to, and dissemination of" science and technology data (CODATA Workshop on Archiving Scientific & Technical (S&T) DATA: report, 20-21 May 2002). This report lists the tasks that are included in data archiving, as: "digitizing data, gathering digitized data into archive collections, describing the collected data to support long-term preservation, decreasing the risks of losing data, and providing easy ways to make the data accessible." In turn, the Federation of Earth Science Information Partners define data archiving as "formally preserving data and information (any type of data or information, e.g. a physical sample, a medieval manuscript, a photograph, or a digital file) and making it available for an identified but potentially large and changing group of data consumers or users" (Preservation definitions, 2011). Finally, Rouse (2010) describes data archiving as "the process of moving data that is no longer actively used to a separate data storage device for long-term retention." Data archives according to her contain older data that are still essential for future reference, as well as data that have to be retained for regulatory compliance. Data archiving, according to her, should not be confused with data backups, which are copies of data. "Data backups are used to restore data in case it is corrupted or destroyed," but data archives, in contrast, "protect older information that is not needed for everyday operations, but may occasionally need to be accessed" (Rouse, 2010).

In summary, data archiving is not just saving or backups of data, but is the process of retention and storage of valuable data for long-term preservation, so that the data will be protected from risk (i.e. loss, or corruption), and will be accessible for future use. Data archiving form an important subset of data management.

4.2.3.5 Data Management

The Data Management Association International (DAMA), focuses on the management of data within an organisation, and uses the term Data Resource Management, which they define as "the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise" (DAMA International, 2014). DAMA International (2007) provides another definition of data management, namely that "data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets."

The above definitions define data within the contexts of an organisation and its assets. The management of research data, however, concerns itself with the management of data within the research context.

4.2.3.6 RDM

A variety of definitions can be found in the literature on the concept of RDM.

Penn State University Libraries define data management in the context of research as “the process of controlling the information generated during a research project.” They see the management of data as “an integral part of the research process,” which “can be challenging particularly when studies involve several researchers and/or when studies are conducted from multiple locations. How data is managed depends on the types of data involved, how data is collected and stored, and how it is used - throughout the research lifecycle” (Penn State University Libraries, 2014).

Texas A&M University Libraries define data management in the context of research and scholarship as “the storage, access and preservation of data produced from a given investigation. Data management practices, according to them, cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long-term preservation of data deliverables after the research investigation has concluded” (Texas A & M University Libraries, n.d.). The
University of Tennessee Libraries, in turn, describe RDM as “the organization and management of data, from its entry into the research lifecycle to the dissemination and archiving of valuable results” (University of Tennessee Libraries, 2014).

The above definitions can be synthesised in the following definition:

> RDM is the process of controlling and organising the data generated during a research project, and covers the entire data lifecycle, which includes the planning of the investigation, conducting the investigation, storage and backing up of the data as it is created, preserving the data long-term, after the research investigation has concluded, and making the data accessible for future use.

4.2.3.7 Critical Summary

The overview of these varying concepts - data curation, data stewardship, data governance, data archiving, and data management with their respective definitions - shows the different nuances of each of these concepts, and how they are sometimes (miss)used as synonymous with RDM, but also how they could be seen as subsets of RDM. The various definitions, characteristics and nuances have been synthesised in Table 4.2.

The approach followed by this study has been to use the concept RDM as an overarching concept, and the other concepts, i.e. data curation, data stewardship, data governance, and data archiving, as subsets. During the process of RDM, value is added to the data through data curation, while someone takes responsibility for the data sets and its tactical function through data stewardship. The process of RDM also includes the governance of data during the research lifecycle, which comprises the goals, policies, shared decision-making, planning, strategies, and processes followed. RDM does not necessarily cover the concept of data management per se, because data management is more focused on data within an organisational context, but could include some of this data if it has research value.

Table 4.2: Concepts, Definitions And Characteristics

Concept	Definition	Characteristics
Data curation	The active and ongoing activities that data stewards engage in to add value to research data throughout its entire lifecycle so that the data are meaningful and useful to scholarship, research and education, and available for discovery and reuse.	• A value-added activity. • Data curation promotes the use of data from its point of creation. • Data curation enables data discovery retrieval and re-use. • Data curation maintains quality. • Data curation adds value, and provides for re-use over time. • Data curation "includes authentication, archiving, management, preservation, retrieval, and

		representation" (Dempsey, 2007). - Data curation is a subset of data management (a layer within data management).
Data stewardship	Data stewardship can be described as a specific approach to data management; it is about taking responsibility for data sets, and is a tactical function that is executed against specific data criteria (Haines, 2012).	- Taking responsibility for data sets. - Data stewardship entails: o Defining the data by identifying key data, gathering definitions, and documenting allowable values; o Defining business rules that can be applied in the creation of data, the usage of data, and derivation of data; o Documenting data sources, for example identifying the system of the record / the system of recommendation; o Setting data quality targets, with fit-for-use thresholds; o Adding metadata for identification / documentation; and; o Remediation of data issues (Haines, 2012).
Data governance	Data governance is concerned with the people managing the data, which includes goals, policies, shared decision making, planning, strategies, and processes followed.	- Data governance is a strategic function. - It focuses on the people managing the data. - It looks at issues such as goals and principles, a communication plan, planning, strategies, policies, processes, shared decision making and roles and responsibilities.
Data archiving	Data archiving is not just saving or undertaking backups of data, but is the process of retention and storage of valuable data for long-term preservation, so that the data will be protected from risk (i.e. loss, or corruption), and will be accessible for future use.	- Data archiving refers to storage and collection of data into archive collections. - It refers to digitisation of data.

		• It refers to long-term preservation of data. • Data archiving protect against the risk of data loss or corruption. • Data archiving makes data accessible. • Data archiving form an important subset of data management.
Data management	Definitions of this concept describe data within the contexts of an organisation and its assets. For example, "data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets" (DAMA International, 2007).	• Data management refers to the management of the full data lifecycle needs of an organisation.
RDM	RDM is the process of controlling and organising the data generated during a research project, and covers the entire data lifecycle, which includes the planning of the investigation, conducting the investigation, storage and backing up of the data as it is created, and preserving the data long-term, after the research investigation has concluded.	• RDM concerns itself with the management of data throughout the whole research process / research lifecycle.

Internationally, RDM as a process is gaining prominence. In the following section, the researcher investigates various international RDM initiatives.

4.3 INTERNATIONAL RDM INITIATIVES

4.3.1 Introduction

Countries around the world are in various stages of developing RDM programmes. Europe (specifically the UK and the European Union), the USA and Australia have taken the lead in RDM, and have the best developed RDM programmes internationally, as can be seen in the literature and websites (UK Data Archive, 2007a: 2-5; Pryor, 2014: 1-8; Mossink, Bijsterbos and Nortier, 2013: 1-10; ESFRI, 2011: 7-8; European Commission, 2017; Data Curation Profiles Toolkit, n.d.; DataOne: Data Observation Network for Earth, n.d.; Purdue University, 2013; California Digital Library, 2014; The National Data Service: a vision for accelerating discovery through data sharing, 2014; Treloar, Choudhury and Michener, 2012: 174). In the discussion on international RDM developments, the researcher of this study decided not to give an exhaustive overview, but rather to focus on the most important country developments as reflected in literature available in English.

4.3.2 UK

RDM initiatives in the UK have also been varied, with some of the largest including the UK Data Archive and the Digital Curation Centre (DCC), and the UK Data Service, funded by the UK government.

4.3.2.1 The UK Data Archive

The UK Data Archive is a national centre for data archiving in the UK. It contains and curates the largest humanities and social sciences digital data collection in the UK (The Royal Society, 2016). Its repository is certified under the Data Seal of Approval (a certification that ensures the safeguarding of data and high quality, and giving guidelines on reliable management of data for the future, without necessitating the application of new standards, regulations or high costs) (Data Seal of Approval, n.d.). The UK Data Archive repository holds thousands of “datasets relating to society, both historical and contemporary” (The Royal Society, 2016). The UK Data Archive renders services to the following UK-based institutions and entities: JISC, the Economic and Social Research Council (ESRC), the Economic and Social Data Service, the Secure Data Service, the Census Registration Service, and the Census Portal (The Royal Society, 2016). The UK Data Archive is funded predominantly by the JISC, the ESRC, and the University of Essex, and is located at the University of Essex (The Royal Society, 2016).

The UK Data Archive has its origin in the SSRC Databank (Social Science Research Council Databank), which was launched at the University of Essex in 1967 with the aim of archiving social and economic research surveys. Since then, the UK Data Archive underwent a number of transformations before it developed into its current form. In 1972, the SSRC Databank was renamed the Survey Archive, which included government surveys from the Government Statistical Service. In 1982, it was again renamed to the SSRC Archive, to signify that a wider range of data resources, other “than just surveys, was now being collected and stored" (UK Data Archive, 2007a: 2-5). Another renaming followed in 1984. This time it was renamed to the Economic and Social Research Council (ESRC) Archive, with a greater focus on empirical research and research considered to be of 'public concern'. Pressures on funding, however, led to less spending on primary data collection and an increase in secondary use of research data, as well as a wider acceptance of data sharing (UK Data Archive, 2007a: 6). The Archive also expanded through its participation in a number of large co-operative data-orientated projects, for example the Rural Areas Database, and the Domesday Project. This set in motion a trend that is still continuing (UK Data Archive, 2007a: 6).

The 1990's was characterised by an expansion of the UK Data Archive, influenced amongst other things by the 1993 UK White Paper on Science and Technology, which emphasised "wealth creation and the need to establish closer and deeper partnership between the academic community and users of its research" (UK Data Archive, 2007a: 6). An example of this expansion was the formation in 1992 of the History Data Unit as a specialist unit within the Archive, becoming part of the Arts and Humanities Data Service (AHDS) three years later and resulting in a renaming of the Unit as the History Data Service (UK Data Archive, 2007a: 7). In 1996, the JISC provided funding to the ESRC Archive in recognition of the support provided by the Archive for teaching and learning. This led to dropping the Council Prefix to the name, resulting in another renaming to become The Data Archive (UK Data Archive, 2007a: 7). In 2000, the final renaming took place when the name changed to the UK Data Archive to indicate both its UK-wide sphere of responsibility, as well as its importance within the international data network (UK Data Archive, 2007a: 7). The participation in projects increased throughout the nineties into the first decade of the new millennium, for example:

- Networked Social Science Tools and Resources (Nesstar) project, focusing on internet developments of value to the European data archives and their users (this included an integrated yet distributed catalogue of data holdings, with additional modules devoted to data browsing, simple analysis, data sub setting and downloading, and data visualisation) (UK Data Archive, 2007b);
- Flexible Access to Statistics, Tables and Electronic Resources (FASTER) project**,** **"**to develop a robust architecture for the dissemination and use of statistics" (UK Data Archive, 2007b);

- Multilingual Access to Data Infrastructures of the European Research Area (MADIERA) project, “to develop and employ a multilingual thesaurus to break down language barriers in the discovery of key social science data resources” (UK Data Archive, 2007b);
- Collection of Historical and Contemporary Census Data (CHCC) project, to develop learning and teaching resources “by improving accessibility to the primary data resources and by developing an integrated set of learning and teaching materials” (UK Data Archive, 2007b);
- Teaching Resources and Materials for Social Scientists (TRAMSS) project, “to place exemplar data analyses in a substantive context by introducing data sources and methods via research questions” (UK Data Archive, 2007b);
- The JISC Exchange for Learning Project (JISC X4L), which fashioned, piloted and recorded “the evaluation of a set of data-based resources for use in teaching in political science courses within higher and further education institutions” (UK Data Archive, 2007b);
- Archive documentation digitisation project, to scan the entire paper documentation holdings of the UK Data Archive and provide machine-readable documentation for all datasets (UK Data Archive, 2007b);
- Online Historical Population Reports (OHPR) project, funded by JISC, that was a significant digitisation undertaking to capture and upload to the web a complete collection of British historical population reports from 1801 to 1937” (UK Data Archive, 2007b);
- The ESRC Qualitative Archiving and Data Sharing Scheme (QUADS), “to develop and promote innovative methodological approaches to the archiving, sharing, re-use and secondary analysis of qualitative research and data, in all of their disparate shape and forms” (UK Data Archive, 2007b);
- The MetaNet network of excellence project that sought to harmonise and synthesise the numerous statistical metadata developments taking place, running from 2000 - 2003 (UK Data Archive, 2016);
- Cluster of Systems of Metadata for Official Statistics (COSMOS), running from 2001-2003 (UK Data Archive, 2016);
- Development of the Collection of Historical and Contemporary Census data and related materials (CHCC) into a major teaching and learning resource, running from 2000 - 2003 (UK Data Archive, 2016);

- Accompanying Measure to Research and Development in Official Statistics (AMRADS), running 2001-2003 (UK Data Archive, 2016);
- GeoXwalk Gazetteer Project (Phases 2 and 3), running 2003-2004 (UK Data Archive, 2016);
- The Geo-Data Portal Project (Go-Geo) (Phases 2 and 3), running 2002-2004 (UK Data Archive, 2016);
- Digital Archives Regional Pilot (DARP), running 2004-2005 (UK Data Archive, 2016);
- Assessment of UK Data Archive and The National Archives (TNA) compliance with Open Archival Information System/Metadata Encoding and Transmission Standard (OAIS/METS), running 2004-2005 (UK Data Archive, 2016);
- Metadata Management and Production System for Surveys in Empirical Socio-economic Research (MetaDater), running 2002-2005 (UK Data Archive, 2016);
- Shibboleth Authentication for Access to the Resource Infrastructures of the UK Data Archive (SAFARI UKDA), running 2005-2006 (UK Data Archive, 2016);
- Census Registration Service (CRS), running 2001-2006 (UK Data Archive, 2016);
- Smart Qualitative Data: Methods and Community Tools for Data Mark-Up (SQUAD), running 2005-2006 (UK Data Archive, 2016);
- TNA Social Science Scoping Study, running January-May 2007 (UK Data Archive, 2016);
- Evaluation of a digital transcription of English parochial registers, 1538-1851: a pilot study, running 2007-2008 (UK Data Archive, 2016);
- Data Exchange Tools and Conversion Utilities (DexT), running 2006-2008 (UK Data Archive, 2016);
- Source to Output Repositories (StORe), running 2005-2008 (UK Data Archive, 2016);
- Preparatory phase project for a major upgrade of the Council of European Social Science Data Archives research infrastructure (CESSDA PPP), running 2008-2010 (UK Data Archive, 2016);
- Rural Economy and Land Use Programme (Relu) Data Support Service, running 2005-2010 (UK Data Archive, 2016);
- Data management planning for ESRC research data-rich investments (DMP-ESRC), running 2010-2011 (UK Data Archive, 2016);

- Semantic technologies for the enhancement of case based learning (ENSEMBLE), running 2008-2011 (UK Data Archive, 2016);
- Census.ac.uk (home of the ESRC Census Programme), running 2006-2012 (UK Data Archive, 2016);
- Secure Data Service, running 2009-2012 (UK Data Archive, 2016);
- The History Data Service (HDS), running 2008-2012 (UK Data Archive, 2016);
- The Survey Resources Network (SRN), running 2008-2012 (UK Data Archive, 2016); and
- Economic and Social Data Service (ESDS), running 2002-2012 (UK Data Archive, 2016).

The second decade of the new millennium saw the roll-out of more projects. These were:

- Managing and curating digital data: Advisory and preservation service, running January-July 2011;
- Using data in sociology teaching, running January-July 2011 (UK Data Archive, 2016);
- Data handling in NVivo 9: online training module, running February-August 2011 (UK Data Archive, 2016);
- Identity management in a service provision environment, running February-August 2011 (UK Data Archive, 2016);
- Researcher Development Initiative (RDI), running 2010-2011 (UK Data Archive, 2016);
- Unlocking the geospatial potential of survey data (UGeo), running January – November 2011 (UK Data Archive, 2016);
- Apply the Simple Knowledge Organization System to the Humanities and Social Science Electronic Thesaurus (SKOS-HASSET), running 2012-2013 (UK Data Archive, 2016);
- Research Data@Essex, running 2011-2013 (UK Data Archive, 2016);
- UK research data registry, running 2012-2014 (UK Data Archive, 2016);
- Enhancing and Enriching Historic Census Microdata, running 2012-2014 (UK Data Archive, 2016);
- Incentives for Data Sharing running February-May 2014 (UK Data Archive, 2016);

- Support for Establishment of National/Regional Social Sciences Data Archives (SERSCIDA), running 2012-2014 (UK Data Archive, 2016);
- Collaboration to Clarify the Costs of Curation (4C), running 2013-2015 (UK Data Archive, 2016);
- Data without Boundaries project (DwB), running 2011-2015 (UK Data Archive, 2016);
- Digital Services Infrastructure for Social Sciences and Humanities (DASISH), running 2012-2015 (UK Data Archive, 2016);
- Alliance for Permanent Access to the Records of Science Network (APARSEN) running 2011-2015 (UK Data Archive, 2016);
- From 2011-2013 the UK Data Archive developed ReCollect, an ePrints plugin to install an institutional data repository pilot for the University of Essex, as part of the Research Data @ Essex project (Van den Eynden et al., 2014); and
- The ReCollect development was followed by ReShare in 2013-2014, a self-deposit subject repository for short-term management curation, developed by the UK Data Archive for the wider social sciences & humanities in the UK, and with funding obtained from the ESRC (Van den Eynden et al., 2014).

The UK Data Archive has also been awarded funding for the following two projects running from 2012-2017:

(a) CESSDA-ELSST

This project aims to align and enhance the Humanities and Social Sciences Electronic Thesaurus (HASSET) and the European Language Social Science Thesaurus (ELSST). These two thesauri support resource discovery for both the UK Data Service and the Consortium of European Social Science Data Archives (CESSDA) (UK Data Archive, 2016).

(b) Cohort And Longitudinal Studies Enhancement Resources (CLOSER)

This project is run in collaboration with the British Library and aims to stimulate interdisciplinary research, develop shared resources, provide training, and share expertise. Central to CLOSER is the development of a metadata discovery platform that provides a portal to hundreds of thousands of variables, questions, and data collection instruments from across the CLOSER portfolio of studies (UK Data Archive, 2016).

In 2012, the UK Data Archive went through an organisational transformation and many of the services that had been hosted by the UK Data Archive were consolidated under the UK Data Service (New national digital repository for social and economic data, 2012a). These were: Economic and Social Data Service (ESDS), History Data Service, Census.ac.uk, Rural Economy and Land Use Programme (RELU) Data Support Service, Secure Data Service, Survey Resources Network, UKDA StatServe, and ESRC Data Store (New national digital repository for social and economic data, 2012a).

From 2015 – 2016, the UK Data Archive also released a number of guides and procedure documents. These are:

- UK Data Service Guidelines for prospective data purchasers;
- UK Data Archive Qualitative Data Ingest Processing Procedures;
- UK Data Archive Documentation Processing Procedure;
- UK Data Archive Quantitative Data Processing Procedures;
- UK Data Archive Data Processing Standard;
- UK Data Archive Data Processing Quick Reference Guide;
- UK Data Service Collections Development Selection and Appraisal Criteria; and
- UK Data Archive Cataloguing Procedures and Guidelines (UK Data Archive, 2016).

In 2016, the UK Data Archive released two policy documents, namely:

- UK Data Service Collections Development Policy, which provides an overview of the selection and appraisal criteria that the UK Data Archive apply to its data collection holdings, in order to ensure the data holdings meet the best needs of all its stakeholders;
- UK Data Archive Preservation Policy, which sets out the UK Data Archive's criteria for the long-term preservation and accessibility of digital objects in its data collections (UK Data Archive, 2016).

4.3.2.2 UK Data Service

The UK Data Service was established in October 2012 by the Economic and Social Research Council (ESRC) (UK Data Service, 2016). Funding for the establishment of the Service primarily came from the ESRC, but further funding was also provided by its host organisations. In 2013, it received additional funding for the coordination of the Administrative Data Research Network (ADRN) for the purpose of streamlining research access to data that are routinely collected by the UK government departments and other entities (Essex receives £5 million for new Big Data Network Centre, 2013). In 2012, a number of services that had been hosted by the UK Data Archive were brought together under the UK Data Service. These were the Economic and Social Data Service, the History Data Service, Census.ac.uk, the Rural Economy and Land-Use Programme (RELU) Data Support Service, the Secure Data Service, Survey Resources Network, the UKDA StatServe, and the ESRC Data Store (New National digital repository for social and economic data, 2012b). In 2013, the UK Data Service also developed a data management costing tool and checklist to assist researchers and institutions in formulating RDM costs in advance, before research commences. This could, for example, be included in a data management plan (DMP) or in the application for funding. The tool takes into account the additional costs that would be needed to preserve research data and make them shareable, while the checklist stipulates activities that should be considered, as well as the costs that would enable effective data management (UK Data Service, 2013).

The UK Data Service is directed and managed by the UK Data Archive, while a Governing Board sees to it that the Service is developed, managed and maintained in a way that would ensure its benefit as a long-term data resource of international value (UK Data Service, 2016). Services in the UK Data Service are provided by staff with special expertise in research data, based at UK host institutions (UK Data Service, 2016). The host organisations are:

(a) Cathie Marsh Institute For Social Research (CMIST)

The CMIST is situated at the University of Manchester, and is a research centre “specialising in the application of advanced quantitative methods in an interdisciplinary social science context" (UK Data Service, 2016). The CMIST is a key role-player in supporting and developing UK micro data of secondary nature, and includes "the Sample of Anonymised Records (SARs) from the census" (UK Data Service, 2016).

(b) Centre For Advanced Spatial Analysis (CASA)

CASA is located at the University College London, and is one of the leading institutions engaged in the science of cities. CASA generates new knowledge and insights that can be applied in city planning, policy and design, and employs the "latest geospatial methods and ideas in computer-based visualisation and modelling" (UK Data Service, 2016).

(c) Department Of Information Studies, University College London (UCL)

As an international centre for knowledge creation in the areas of librarianship, information science, archives, records management, publishing, and digital humanities, the UCL Department of Information Science brings together researchers and practitioners in these fields, in order to gain insights and understanding that would be needed to frame the emerging information environment. At the same time, the department is unravelling and building on the historical developments that have formed this environment (UK Data Service, 2016).

(d) EDINA

EDINA is an academic centre that provides digital expertise nationally in the UK, and internationally, and is located at the University of Edinburgh (EDINA, n.d.; UK Data Service, 2016). EDINA started operating in 1995 and has been assigned by JISC to support developments at universities, colleges, and research institutes in the UK. Its mission is to develop and deliver shared services and infrastructure for research and education, which are innovative and of high quality and cost effective (UK Data Service, 2016). EDINA engage with emerging technologies and transforms "new ideas from prototypes, through research and development to scalable digital solutions" (EDINA, n.d.). Examples of this include "geospatial services, crowdsourcing tools, cultural resources, and pioneering mobile apps" (EDINA, n.d.). EDINA is also operating as the

ESPRC specialist geography unit for the UK census, and supports users of the UK Data Service by providing access to the geography outputs of the 2011 and earlier censuses (UK Data Service, 2016).

(e) Population, Health And Wellbeing Research Group In Geography And Environment

This group, located at the University of Southampton, is well known in the UK for leadership in population and health research, using a blend of innovation in methodology that involves "geographical information systems (GIS), spatial analysis and quantitative and qualitative methods" (UK Data Service, 2016). "Spatial population analysis and modelling" as well as "cultures, spaces and practices of care and population health" are the core issues the group deals with (UK Data Service, 2016). The group conducts its work in collaboration with the UK Office for National Statistics, co-directs the ESRC National Centre for Research Methods and UK Data Service, and in addition, edits the journal *Health and Place* (UK Data Service, 2016).

(f) School Of Geography, University Of Leeds.

The School is known globally as one of the leading geography departments, and recognised for the cutting edge-research that is done there. Its research has an impact on policy, covering a wide array of sectors. The School also houses the Centre for Spatial Analysis and Policy, where imaginative analysis techniques and policy predictions are developed. Furthermore, the School has been associated with both the spatial analysis of census data, as well as the "development of web-based systems" that could be used in the broader academic community (UK Data Service, 2016).

(g) UK Data Archive

For more information, see 4.3.2.5.

(h) JISC

The JISC is a not-for-profit organisation situated in the UK, that furthers the use of digital services and technologies among the higher education, further education and skills sectors of the UK (UK Data Service, 2016; JISC, 2016a). For more on the RDM programmes of the JISC, see 4.3.2.4. The UK Data Service forms part of the Digital Resources Directorate of JISC, and UK Data Service personnel gives access to, and renders specialist support for, the databanks of inter-governmental organisations, for example the World Bank, the Organisation for Economic Co-operation and Development (OECD), and the International Monetary Fund (IMF). The UK Data Service also provides access to and support for “aggregate statistics from the 1971 to 2011 UK Censuses” (UK Data Service, 2016).

4.3.2.3 The Digital Curation Centre (DCC)

The DCC is a UK-supported service and centre of excellence for digital preservation and data management (Donnelly, 2013: 37). The DCC was preceded by an e-Science initiative introduced in the UK governments’ spending review of 2000, aimed at encouraging the development of an IT infrastructure that would be adequate to support the increasingly international research collaborations developing from science and engineering disciplines (Pryor, 2014: 1). In 2003, the JISC published its ‘Circular 6/03 (Revised) Digital Curation Centre’. The circular invited proposals for the establishment of a National Digital Curation Centre for the UK that would take the “lead in research and development into key areas of digital curation for data and publications” (Circular 6/03 (Revised) Digital Curation Centre, 2012). It also proposed that such a centre should “pilot the development of generic support services for maintaining digital data and research results over their entire life-cycle for current and future users” (Circular 6/03 (Revised) Digital Curation Centre, 2012). Circular 6/03 also determined that data curation “includes all the processes needed for good data creation and management, and the capacity to add value to generate new sources of information and knowledge,” which infers that there must be a sustained interaction between creators, suppliers, archivers and consumers of data (Pryor, 2014: 2). On 1 March 2004, the DCC was officially launched, following a successful response to JISC Circular 6/03 by a consortium comprising the Universities of Edinburgh and Glasgow (which together hosted the National e-Science Centre), UKOLN at the University of Bath, and STFC, which managed the Rutherford Appleton and Daresbury Laboratories (DCC, 2016).

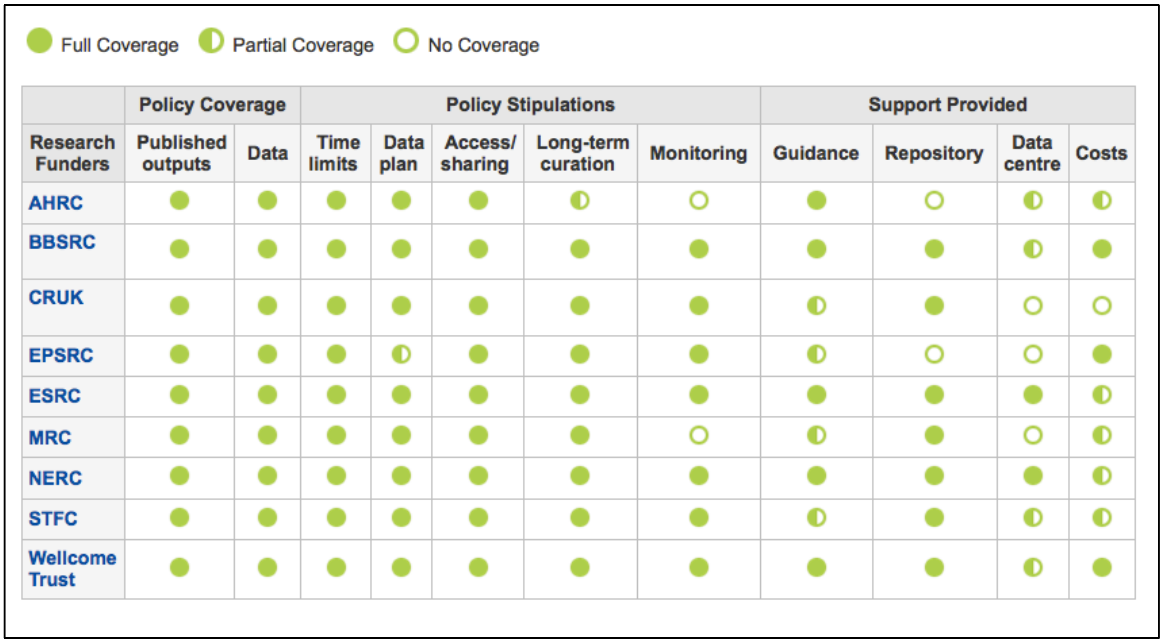
During Phase 1 (March 2004-February 2007) and Phase 2 (March 2007-February 2007) of the roll-out of the DCC, people involved in digital preservation and curation activities within higher and further education in the UK were targeted. This included data specialists, policy-makers and information professionals such as librarians, archivists, records managers, as well as researchers (those that create the data) (DCC, 2016). During these two phases, the DC also reached out to the public and commercial sectors, sister organisations, and standards workings groups, because it recognised that the development of tools and processes for digital curation lay beyond the UK higher education and further education sector as well as within it (DCC, 2016). This, in turn, led to the establishment of the "DCC Associates Network as a forum for cross-sectoral communication on important problems" in digital curation (DCC, 2016). By the time Phase 2 commenced, the focus of the DCC activity had shifted to a large extent towards a heightened and direct involvement with the active research community (DCC, 2016).

In Phase 3 (March 2010-February 2013), further structural changes were introduced, which characterised a move away from the development of curation tools and a renewed emphasis on building capacity, capability and development of skills for data curation throughout the higher education research community in the UK (DCC, 2016). During this time frame, the DCC also undertook a significant outreach programme to assist a selected group of universities in the development of their RDM capabilities (Donnelly, 2013: 38). The programme was funded through the HEFCE's Universities Modernisation Fund and became known as the Institutional Engagement (IE) Programme (Donnelly, 2013: 38). Engagements were tailored to fit the specific needs of each institution. The focus in each engagement was on research practitioners; support staff from research offices, libraries and IT Departments; and senior managers that were responsible for university budgets (Donnelly, 2013, 39). The engagements consisted of actions that included the development of RDM roadmaps and policies, identification of training and support needs, and trialling and customisation of the Data Asset Framework (DAF), a tool that "provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets," DMPOnline, the DCC's data management planning tool, and the Collaborative Assessment of Research Data

Infrastructure and Objectives (CARDIO), with the aim to integrate these with an institution's existing technical infrastructures (DCC, 2017; DCC, 2014d; Donnelly, 2013: 39).

The DCC currently provides a wide variety of valuable practical resources/guides and expertise on RDM, for example the DMPonline "tool that assists researchers to produce an effective DMP to cater for the whole lifecycle of a project, from bid-preparation stage through to completion" (DCC, 2014d). Another helpful tool is the research that was done on comparing the various funders' requirements for RDM and tabling these (DCC, 2014a) (See Figure 4.2).

Figure 4.2: Funder Requirements: UK (DCC, 2014a)



● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

Early in 2016, the DCC and the California Digital Library (CDL) joined forces in writing a proposal for the building of a new global data management advisory platform, Roadmap, that would include the work done by the DCC in DMPonline and the work done by CDL on the DMPTool, integrated with other components of the research lifecycle (Simms et al., 2016: 1). This proposal contemplated closer engagement with individual disciplinary communities. DMPonline, primarily supporting researchers in the UK, and the DMPTool, primarily supporting researchers in the USA, were described as first-generation service offerings, which assisted researchers in "fulfilling their data management planning obligations" as set out in "funder mandates, pre-publication requirements, and institutional policies" (Simms et al., 2016: 3). The idea behind the Roadmap is to converge these tools into “a common technical platform” that would offer RDM advice and would underpin open science (Simms, et al., 2016: 3). The new platform would integrate all the existing functionalities of both the UK and USA tools. It would also “reposition the DMPs as living documents” that could be useful in reconfiguring the flow of research activities, as well as integrate with associated “data management systems and workflows,” for example the Open Science Framework, SHARE (an independent volunteer-run information technology association that provides education, professional networking and industry influence), the Crossref / Datacite DOI Event Tracking System and Zenodo (Roadmap: global research data management advisory platform combines DMPTool and DMPOnline, 2016; Simms, et al., 2016: 3).

At the time of writing this thesis, the plan was to roll out the project in two phases. Phase I contained the process of delineating their co-development process and partnership agreement, starting with a gap analysis of the two tools and the consolidation of the roadmap. New features were to be added in the second half of 2016 and the coordinated release was expected in the first half of 2017. These features included:

- Extension of authentication and localisation “support for all instances”;
- Identification of partners and provision “of an integration roadmap for external/reporting systems”; and
- Restructuring the concept of themes in DMPonline into an actionable data model for funder prerequisites (Simms, et al., 2016: 6).

As part of the above process, the DCC and the CDL plan to maintain outreach and training programmes nationally and continue with international outreach efforts (Simms, et al., 2016: 6). They foresee that this will be in the form of meetings with funders and researchers to evaluate prevalent practices and workflows and establish “additional points of integration with existing systems, metadata requirements, etc.” (Simms et al., 2016: 6). The planning of these meetings would be done in consultation with OpenAIRE, ELIXIR, EUDAT, bioCADDIE and BioSharing (Simms, et al., 2016: 6). In Phase II, the plan is to proceed “with a second coordinated product release that would” encompass additional components “and integrations, and pursue additional funder/researcher events” (Simms, et al., 2016: 6).

4.3.2.4 JISC

As mentioned in 4.3.2.3, JISC furthers the use of digital services and technologies among the higher education, further education and skills sectors of the UK by providing funding. As part of this service, JISC released its first Managing Research Data (JISCMRD) programme in October 2009. This programme was completed in September 2011 (JISC, 2014a). The programme targeted a number of key areas:

- Piloting crucial RDM infrastructures within institutions and for distributed research teams;
- Improving the practices of data management planning;
- Development of tools to assist institutions in the planning of their RDM practices;
- Promoting the publishing of research data and demonstrating the advantages of enhanced procedures for citing, linking and integrating data; and
- Encouraging the acquisition of suitable skills among researchers and research support personnel in universities (JISC, 2014a).

The JISCMRD programme comprised five strands:

- RDM Infrastructure (RDMI) Projects;
- RDM Planning (RDMP) Projects;
- Support and Tools Projects;
- Citing, Linking, Integrating and Publishing Research data (CLIP) Projects; and
- Research Data Management Training Materials Projects (JISC, 2014a).

In October 2011, JISC released its second Managing Research Data Programme, which ran until July 2013 (JISC, 2014a). The aim with this second programme was to build on work that was done in the first programme. This was done by broadening the implementation base of the innovations that flowed from the first programme, as well as enhancing the infrastructure and practices (JISC, 2014a). The programme also generated a number of valuable software supporting systems, guidance, as well as policies that could be used by other institutions (JISC, 2014a). The programme comprised of the following:

- Seventeen large institutional projects assisting universities to pilot or further develop and expand infrastructures for RDM;
- Eight projects assisting research groups, projects or departments to "fulfil disciplinary best practice," as well as "the requirements of research funders by implementing data management plans and supporting systems"; and
- Two projects that focused on customising the DCCs DMPOnline tool for institutional tools (JISC, 2014a).

From July 2014 to July 2016, JISC ran a project called Research at Risk. In this project, they worked in partnership with UK universities, relevant professional associations and national organisations such as research councils, to provide infrastructure, advice and tools that would support UK universities and their researchers in establishing successful data management practices as part of their core research activities (JISC, n.d.). The work of this project included the following:

- Establishing the right policies and actions that would undergird the "management and use of research data", as well as "to respond to funder mandates";
- Developing guidance that would assist universities in their response to Research Council policies;
- Providing templates for the development of "a RDM business case and related templates for costings";
- Establishing "a research data discovery service for the UK";
- Developing "preservation and storage services," for example "the Arkivum framework agreement";
- Instigating key standards that would underpin "more effective and efficient management of research data and research information", encompassing ORCID;
- Developing "experiments and prototypes for new solutions to keep RDM up to date," for example the Research Data Spring project and competition; and
- Supporting "the development of skills and capabilities of research data managers" (which included researchers, research managers, and library and IT staff) (JISC, n.d.).

A key output of the Research at Risk project was the report on 'Directions for Research Data Management in UK Universities,' which was published in March 2015. This report was published by JISC in collaboration with the Association of Research Managers and

Administrators (ARMA), Research Libraries UK (RLUK), the Russell Universities Group of IT Directors (RUGIT), the Society of College, National and University Libraries (SCONUL), and the Universities and Colleges Information Systems Association (UCISA) (Brown, Bruce and Kernohan, 2015: 4). This report discussed a collaborative direction for RDM for universities in the UK over the next five years and covered five key areas that would require action at national and institutional level. These are:

- "Policy development and implementation
- Skills and capability;
- Infrastructure and interoperability;
- Incentives for researchers and support stakeholders; and
- Business case and sustainability" (Brown, Bruce and Kernohan, 2015: 4, 6).

Each of the above-mentioned areas were elaborated upon as follows:

(a) Policy Development And Implementation

The report showed that more work with funders is needed to assist universities in understanding funder requirements better (Brown, Bruce and Kernohan, 2015: 7). It also revealed that ways would need to be found to instil a reward culture that would motivate researchers to get involved willingly with RDM processes (Brown, Bruce and Kernohan, 2015: 7). The report further showed that relatively few universities had, at that time, "adopted an RDM policy" (Brown, Bruce and Kernohan, 2015: 7). With regards to policy development, the results also varied. Some had created their own policies, while others mimicked or adapted policies of other institutions (Brown, Bruce and Kernohan, 2015: 7). Some of the policies were shown to be focusing on best research practice, while others were merely focusing on complying with funders' requirements (Brown, Bruce and Kernohan, 2015: 8). The report further disclosed that "a 'one size fits all' approach to policy development" would "not be the best course of action" for the future, and that the drivers influencing institutions to develop policies would likely increase and become more insistent (Brown, Bruce and Kernohan, 2015: 8).

(b) Skills And Capabilities

The report divulged that the successful development and implementation of RDM policies will rely on a wide array of skills, but no one is likely to have all the skills that are necessary. The RDM manager would nevertheless need a thorough understanding of all the issues, and a solid set of soft skills to assist in ensuring that complex projects are implemented (Brown, Bruce and Kernohan, 2015: 12). Learning on the job was found to be common, while “training courses by external providers” were viewed “as both useful and effective” (Brown, Bruce and Kernohan, 2015: 12). The report also mentioned that opportunities for research data managers “to shadow their more experienced counterparts could be a useful alternative to formal training” (Brown, Bruce and Kernohan, 2015: 12).

(c) Infrastructure And Interoperability

With regards to infrastructure and interoperability, the report stated that “the interoperation of different systems is desirable” and that the embracing of common metadata standards plays a key role in achieving it (Brown, Bruce and Kernohan, 2015: 16). The report further revealed that there is enthusiasm in the research community for JISC to take “the lead in supporting a common metadata standard for RDM” (Brown, Bruce and Kernohan, 2015: 16). The report also disclosed that disparate institutions are starting from different departure points, where some already have more than sufficient data storage capacity, while others are starting from scratch by acquiring the best-estimate solutions (Brown, Bruce and Kernohan, 2015: 16). In addition, the report revealed that “research-intensive universities” are not likely “to outsource their requirements”, but this might be an attractive alternative for others. Overall, there seemed to be “strong support for shared service solutions” (Brown, Bruce and Kernohan, 2015: 16). Furthermore, the report showed many institutions would be “offering a high-level, basic service to researchers”, whilst not taking into consideration “disciplinary differences in metadata collection” and expecting their researchers to drive the dataset description process themselves (Brown, Bruce and Kernohan, 2015: 16). The report also divulged that approaches varied among institutions in capturing and exposing appropriate metadata, where some were using their Current Research Information System (CRIS) and others built their own systems (Brown, Bruce and

Kernohan, 2015: 16). A further revelation from the report was that researchers are often allocated a specific amount of data storage, but also have the option to put their allocation in a pool with their colleagues (Brown, Bruce and Kernohan, 2015: 12). The need for additional storage was usually met through research grants (Brown, Bruce and Kernohan, 2015: 16). The report finally showed that the storage of data during the active research phase should be addressed, so that collaborations do not suffer (Brown, Bruce and Kernohan, 2015: 16). In addition, institutions would need more mature information management policies, which should correspond with broader work on cybersecurity (Brown, Bruce and Kernohan, 2015: 16).

(d) Incentives For Researchers And Support Stakeholders

The report revealed that compliance to RDM policy will not precipitate researchers to embrace RDM (Brown, Bruce and Kernohan, 2015: 18). In addition, the report stated that it was difficult to convince researchers of the benefits of RDM and long-term storage, and that there were few incentives for them to get involved (Brown, Bruce and Kernohan, 2015: 18). Where there has been an authoritarian approach, researchers' response had been to do just the barest minimum with regards to provision of metadata (Brown, Bruce and Kernohan, 2015: 18). Furthermore, the report revealed that although funders mandate archiving of datasets, researchers feared that costing it into their research proposals would make them uncompetitive (Brown, Bruce and Kernohan, 2015: 18). At the same time, it revealed that "it would be useful to achieve greater clarity about what RDM-related costs" could be "recovered from funders' grants" (Brown, Bruce and Kernohan, 2015: 18).

The report furthermore disclosed that researchers need explicit and meaningful rewards for engaging actively in RDM, while the reward structures at the time of the report were sometimes seen as too focused on publishing in high impact journals (Brown, Bruce and Kernohan, 2015: 18). In addition, "a greater focus on the value that effective RDM" could effect on the publication process would be helpful, and this could be made possible by the availability of more data-focused journals and relevant metrics (Brown, Bruce and Kernohan, 2015: 18). The report further advocated highlighting the benefits of the opening of access to data to the wider research community and society, and stated that making data more shareable would stimulate a cultural change where the re-using of others' "data becomes more common in more disciplines" (Brown, Bruce and Kernohan, 2015: 18). The report also suggested that the possibility of providing "download information and other statistics" would motivate researchers to engage with RDM (Brown, Bruce and Kernohan, 2015: 18). In addition, the report proposed the use of 'data fellows' to coach researchers in publishing data and building collaborative networks. These data fellows should also get career-related rewards for such coaching (Brown, Bruce and Kernohan, 2015: 18). The report further suggested the running of an RDM pilot project that would bring institutional resources to the fore, and also emphasized the importance of the role that librarians could play (Brown, Bruce and Kernohan, 2015: 19). Incentives could be the satisfaction of seeing the institution "comply with external requirements" and "foreseeing and forestalling problems such as data protection issues and information security" (Brown, Bruce and Kernohan, 2015: 19). The report finally stated that local awards could be given to RDM managers (Brown, Bruce and Kernohan, 2015: 19).

(e) Business Case And Sustainability

The report disclosed that "approaches to funding RDM services and infrastructure varied" vastly (Brown, Bruce and Kernohan, 2015: 22). Another aspect that the report highlighted was the general uncertainty that existed about storage capacity that was required at the time and in the future (Brown, Bruce and Kernohan, 2015: 22). The uncertainty "about how much of the cost of RDM" services and infrastructure would be recoverable from funders, as well as the apprehension about the complexity of the process, were also highlighted (Brown, Bruce and Kernohan, 2015: 22). The sustainability of all aspects of RDM was further mentioned as something that will still need to be considered (Brown, Bruce and Kernohan, 2015: 22). Finally, the report mentioned that "good information management should help senior university management to justify" the funding of "RDM infrastructure and services" (Brown, Bruce and Kernohan, 2015: 22).

Another outflow from the Research at Risk project was the development of a Research Data Discovery Service for the UK. This project was scheduled to be concluded at the end of September 2017 (JISC, n.d.). This project was undertaken in three distinct phases. Phase 1 consisted of an initial pilot funded by JISC that ran from October 2013 to March 2014 (JISC, n.d.). In this phase, the DCC and the UK Data Service "piloted an approach to a registry service" that aggregated metadata for research data that were kept in "UK higher education institutions and national, discipline-specific data centres" (JISC, n.d.). Phase 2 comprised the development of the alpha service, and ran from March 2015 to September 2016 (JISC, n.d.). This phase was also supported by the DCC and the UK Data Service, and the aim was to build on the work done in the pilot, by running a test UK Research Discovery Service (JISC, n.d.). Phase 3 ran from October 2016 to September 2017 and focused on moving from the alpha service to an enhanced beta service, in order to deliver a fully functional Research Discovery Service (JISC, n.d.).

Yet another outflow from the Research at Risk project was the Research Data Spring project. This project commenced in October 2014 and was concluded in October 2016. It focused on finding "new technical tools, software and service solutions" that would "improve researchers' workflows and the use and management of their data" (JISC, n.d.). During the first phase of the project, ideas were gathered and an attempt was made to create new connections openly on the web (JISC, n.d.). These ideas were then developed during a sandpit workshop, which ran from 26-27 February 2015, and were subsequently presented to an expert panel (JISC, n.d.). In the second phase, selected ideas were further funded and developed during a three-month period and again presented on 13-14 July 2015, for a third phase of development (JIS, n.d.). In December 2015, another workshop was held and seven projects were selected for the third stage of development, which ended in August 2016 (JISC, n.d.).

Another outcome of the Research at Risk project was the Data Archiving Framework (JISC, n.d.). This framework is described as "a highly secure and easy-to-use and cost-effective archiving service for research and education." It uses Arkivum as supplier following a procurement through the EU (JISC, n.d.). The Framework saves institutions additional costs that could be "incurred by procuring the service directly" and provides "increased levels of compliance," while at the same time reducing spending on IT and administrative costs that goes hand-in-hand with in-house archiving (JISC, n.d.). The Framework can be used by all Janet-connected organisations (Janet is a high-speed network for the UK research and education community) (JISC, n.d.). Finally, as part of the co-design challenge followed through the Research at Risk project, JISC released the Next Generation Research Environments Discovery Phase Report in May 2017. For more on Next Generation Research Environments, see VRE projects, 3.2.1.

4.3.2.5 Other Significant Developments In The UK

In 2009, the Panton Principles were compiled by a group of researchers that were of the opinion that science could only function effectively, and that society will only glean the full value from scientific endeavours, if science data are made open access (Murray-Rust et al., 2010). The Open Knowledge Foundation Working Group on Open Data in Science further refined these principles and introduced it to a wider public platform (Murray-Rust et al., 2010; Pryor, 2014: 4). This was followed by the release of a revised Research Data Policy in September 2010 by the Economic and Social Research Council (ESRC), which in turn led to a new requirement in the Spring of 2011, that obligated research grant applicants to submit a statement on data sharing and hand in a data management and sharing plan together with their applications (Horton et al., 2011: 3). Support to researchers was provided by the ESRC's UK Data Archive service staff. It included assistance to researchers in planning their data management and sharing, and continuous aid during the course of their project, including the final deposit and re-use of data (Pryor, 2014: 7).

The next development in the UK came in 2011, when the 'Common Principles on Data Policy', although not enforceable, was published by Research Councils UK (RCUK). These principles offered a comprehensive framework for individual UK "Research Council policies on data policy" (Research Councils UK, 2014). The document included a principle that declared publicly funded research data as being a public good, "produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property" (Research Councils UK, 2014). The document assumed that research institutions would have data policies and plans in place, and that actions would be taken to preserve data of value along with appropriate metadata to make the data understandable, retrievable and re-useable by other researchers (Pryor, 2014: 7).

In the same year (2011), the Engineering and Physical Sciences Research Council (EPSRC) published its 'Policy Framework on Research Data', consisting of seven core principles, directly aligned to the RCUK principles (EPSRC, 2013). Two of these principles were specifically emphasized. Firstly, “that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner”; and secondly, “that the research process should not be damaged by the inappropriate release of such data” (EPSRC, 2013). These principles assumed that processes and systems were in place at institutions to enable these actions (Pryor, 2014: 7). The EPSRC also did not provide any support infrastructure, which meant that universities that received EPSRC grants were expected to provide the necessary services and support infrastructure themselves (Pryor, 2014: 7- 8).

In 2012, the Royal Society Science Policy Centre Report 02/12, titled ‘Science as an Open Enterprise: open data for open science’, appeared. This report emphasized the “rapid and pervasive technological change” that “has created new ways of enquiring, storing, manipulating” and transferring huge quantities of data; which in turn encouraged new practices of communication and collaboration among researchers and challenged several norms of established scientific behaviour (Pryor, 2014: 2; The Royal Society, 2012: 3). This report suggested six further changes in research practice that would be required if these new technologies and collaborations are to be exploited fully:

- Move away from a research culture that regards data as a private reserve;
- Expand the criteria used to evaluate research, so that recognition can be granted for beneficial data communication and innovative methods of collaborating;
- Create shared standards for communicating data;
- Mandate “intelligent openness for that data which is” pertinent to published scientific papers;
- Establish “a strong cohort of data scientists to manage and support the use of digital data”; and
- Develop and utilise “new software tools to automate and simplify” the construction and utilisation of datasets (Pryor, 2014: 2-3; The Royal Society, 2012: 3).

In March 2015, the ESRC revised their research data policy further, and their policy are now underpinned by nine core principles, which are aligned with the RCUK Common Principles on Data Policy (ESRC, 2017, UK Data Service, 2016). Guidelines are also provided for their implementation, and “the specific roles and responsibilities of researchers, institutions, the ESRC and its data service providers” are clearly stipulated

(UK Data Service, 2016). Grant holders are also required to use the UK Data Archive as a place to deposit research data, via the UK Data Service ReShare repository (UK Data Service, 2016).

In July 2016, a multi-stakeholder group from the research community in the UK published the Concordat on Open Research Data, with the specific aim "to ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical, disciplinary and regulatory frameworks and norms, and with due regard to the costs involved" (Concordat Working Group, 2016). The Concordat proposed ten principles (expectations of best practice) for working with research data, which include the multiple roles needed to facilitate the research process. These principles as listed by Concordat Working Group (2016) as:

Principle 1: "Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice."

Principle 2: "There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable."

Principle 3: "Open access to research data carries a significant cost, which should be respected by all parties."

Principle 4: "The right of the creators of research data to reasonable first use is recognized."

Principle 5: "Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement."

Principle 6: "Good data management is fundamental to all stages of the research process and should be established at the outset."

Principle 7: "Data curation is vital to make data useful for others and for long-term preservation of data."

Principle 8: "Data supporting publications should be accessible by the publication date and should be in a citeable form."

Principle 9: "Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders."

Principle 10: "Regular reviews of progress towards open research data should be undertaken."

A study conducted by Cox et al. (2017) titled 'Developments in RDM in academic libraries: towards an understanding of research data service maturity', revealed that most institutions in the UK (86%) reported that they do have a RDM policy, and that the EPSRC had been influential in steering institutions towards developing a RDM policy. In some institutions, it also disclosed an increasing maturity in RDM, where RDM was just one element of a broader RDM strategy or roadmap (Cox et al., 2017: 2186). 44% of libraries in the UK were also shown to have frequently led or been involved in evaluative methods such as audit tools or surveys to obtain a better understanding of RDM at their institutions (Cox et al., 2017: 2187).

Cox et al. (2017: 2187) reported funding resources as one of the most challenging aspects of RDM. Funding were largely from financial resources that were not fixed term (Cox et al., 2017: 2187). There was recognition from most countries, including the UK, that funding for RDM would "need to come from multiple sources" (Cox et al., 2017: 2187). For example, "infrastructure funding should, at least in part, be allocated at supra-institutional level," in other words, national top-sliced funding should go into the development of national services to support institutions (Cox et al., 2017: 2187). At institutional level, a number of libraries mentioned that resourcing was problematic, as there were no additional funding resources available than the library budget. They did, however, acknowledge that some staff resources might need to be re-allocated to RDM, with corresponding reallocation and repurposing of funding resources in the personnel budget (Cox et al., 2017: 2188).

The study by Cox et al. (2017: 2188) also revealed the need for intra-institutional collaboration with entities such as IT Services, the research office, faculties, and other academic committees for successful establishment of RDM services at institutions. The study furthermore showed that there was relatively little (22%) collaboration pertaining to RDM with external organisations (Cox et al., 2017: 2188). The majority of responses from UK libraries indicated that they had a basic research data advisory service in place, with 26% indicating that they maintained a web resource guide and 62% indicating that they conducted RDM training and/or data literacy training (Cox et al., 2017: 2189). 74%

indicated that they did not have an advisory service on data analysis, data mining or data visualization (Cox et al., 2017: 2188). The majority of libraries (62%) further indicated that their members were not directly participating with researchers on research projects as team members (Cox et al., 2017: 2190). Most of these libraries signified that they see advisory services, training and RDM website development, as important strategic priorities (Cox et al., 2017: 2190). They also designated the following services as priorities for future development:

- Advisory services for data analysis / data visualisation;
- Search / retrieval of external data sources; and
- Project participation (Cox et al., 2017: 2190).

With regard to technical RDM services, 14% of UK libraries were revealed to offer advisory services on copyright and/or intellectual property issues, and/or licensing property rights relating to data and RDM, while only 14% of libraries were shown to run a data repository/archive store (Cox et al., 2017: 2191). Service areas that were identified as top strategic priorities for future development were the development of advisory services for copyright and intellectual property issues, running of data repositories, development of a data catalogue, and curation of active data (Cox et al., 2017: 2191).

A major challenge that was identified by the majority of institutions was finding library staff with necessary RDM skills. A number of areas were then identified where skills development would be needed, namely data curation, legal policy and advisory skills, data description and documentation (metadata skills), and research methods (Cox et al., 2017: 2191).

4.3.3 The European Union (EU)

Discussions on RDM and the development of RDM in the EU have mostly focused on infrastructure, and specifically how to construct good e-infrastructures for research data, as well as how to set mechanisms in motion among researchers for sharing data (Johnsson and Åhlfeldt, 2015: 13).

4.3.3.1 Consortium Of European Social Science Data Archives (CESSDA)

The Consortium of European Social Science Data Archives (CESSDA) renders wide-ranging, integrated and sustainable data services to the social sciences field (CESSDA, 2016). CESSDA developed from a network of European data service providers into a legal entity with large-scale infrastructure under the authority of ESFRI (CESSDA, 2016). In 2016, CESSDA was recognised as an ESFRI Landmark in the ESFRI 2016 Roadmap (ESFRI, 2016). The ministry of research or delegated institutions of individual member states own and finance CESSDA, while Norway hosts it. Its main functions are:

- "Coordination of the Network of European data service providers and the promotion of the results of social sciences";
- The facilitation of researcher access to key resources that are relevant to the "European social science" programme regardless of where the data or the researcher is located;
- Actively and persistently working to add more resources from Europe and further afield, into the infrastructure;
- Offering of training opportunities within CESSDA and further afield on best practices in RDM and operations;
- Promotion and facilitation of more extensive participation in CESSDA; and
- Creation and advancement, as well as "coordination of standards, protocols and professional best practices" relating to the preservation and distribution of data and digital objects that are linked to these (CESSDA, 2016).

4.3.3.2 The European Strategy Forum On Research Infrastructures (ESFRI)

The European Strategy Forum on Research Infrastructures (ESFRI) was launched in 2002 as an informal forum by the EU Council, and reaffirmed in November 2004, May 2007 and December 2012. ESFRI was set up to:

- Promote a comprehensible and "strategy-led approach to policy making on research infrastructures in Europe";
- "Facilitate multilateral initiatives" resulting in an improved utilization and development of research infrastructures, while performing the role of "incubator for pan-European and global research infrastructures"; and

- Institute “a European Roadmap for research infrastructures” (novel and extensive upgrades, pan-European interest) for the next 10-20 years, encourage the establishment of these facilities, and amend the Roadmap as needs emerge; ensure that the implementation of current ongoing ESFRI projects is followed up after an extensive assessment, as well as the “prioritisation of the infrastructure projects listed in the ESFRI Roadmap” (ESFRI, 2016: 10).

The European Strategy Forum on Research Infrastructures (ESFRI) has focused on integrating and opening national research facilities and developing e-infrastructures underpinning a digital European Research Area, as part of Horizon 2020, the EU Framework Programme for Research and Innovation (ESFRI, 2011: 7-8).

In 2006, ESFRI published its first roadmap for the construction and development of what they termed the “next generation of pan-European research infrastructures.” This was followed by a roadmap in 2008, and another in 2010. The roadmap in 2010 contained 48 projects focusing on the fostering of European leadership across a wide “range of scientific fields” (ESFRI, 2011: 7-8). By 2015, 60% of these ESFRI projects had been completed, and 29 of these had reached the implementation phase. These are currently “pan-European hubs of scientific excellence, generating new ideas and pushing the boundaries of science and technology” (ESFRI, 2016: foreword).

The ESFRI Strategy Report on Research Infrastructures: Roadmap 2016, contained twenty-one ESFRI projects, of which nine came from the 2008 Roadmap, six from the 2010 Roadmap and five were new projects, plus one reoriented project that was selected from twenty project proposals in March 2015 (ESFRI, 2016). These new projects were selected after an evaluation by the Strategy Working Groups of “their scientific excellence, pan-European relevance and socio-economic impact,” as well as their level of maturity as measured against an ‘assessment matrix’ that was “developed by the ESFRI Implementation Group (IG)” (ESFRI, 2016: 13). The ESFRI Strategy Report on Research Infrastructures: Roadmap 2016 also contained twenty-nine ESFRI Landmarks. These are ESFRI projects that have been successfully implemented and were rendering services, or have effectively advanced in their development (ESFRI, 2016: 13).

4.3.3.3 EUDAT

In 2011, the European Commission (EC) and European member states created EUDAT as a pan-European e-infrastructure supporting multiple research communities, resulting in a European e-infrastructure ecosystem, with communication networks, distributed grids and HPC facilities (Mossink, Bijsterbos and Nortier, 2013: 5; Michelini and Lecarpentier, 2011). EUDAT has its origins in the work of the PARADE (Partnership for Accessing Data in Europe) initiative and the PARADE White Paper, which was published in October 2009, titled ‘Strategy for a European Data Infrastructure that should be persistent, multidisciplinary, and based on the need of user communities’ (PARADE, 2009). The work of PARADE was supported and further elaborated upon by a number of policy and expert bodies. In September 2010, the e-Infrastructure Reflection Group (e-IRG) and ESFRI published an e-IRG Blue Paper, which recommended the identification and promotion of common (long-term) data related services across different research infrastructures. This was followed in October 2010 by the “High Level Expert Group (HLEG) report on Scientific Data Infrastructure for scientific data, which supports seamless access, use, re-use and trust of data” (Michelini and Lecarpentier, 2011). This was then followed by the launching of EUDAT in October 2011 (Michelini and Lecarpentier, 2011).

EUDAT provides “common data services” through a geographically distributed network of thirty-five European institutions (Donnelly, 2015). Services are shared and resources are stored across fifteen European countries, while data are stored alongside a number of Europe’s powerful supercomputers (Donnelly, 2015). EUDAT’s vision is to make it possible for “European researchers to preserve, find, access and process data in a trusted environment” that is “part of a Collaborative Data Infrastructure (CDI)”; in other words, a network consisting of collaborating centres, community-specific repositories, and “some of Europe’s largest scientific centres” (Donelly, 2015, EUDAT, n.d.). The mission of EUDAT is to “design, develop, implement and offer Common Data Services” as presented in the ‘Riding the Wave’ report (published in 2010) “to all interested researchers and research communities” (Research Data Alliance, n.d.). Currently, JISC and the DCC are partners, but it is uncertain what the impact of BREXIT (the UK leaving the EU) will have on research partnerships between the UK and the EU (Donnelly, 2015). This might be a topic for future research.

4.3.3.4 European Cloud Initiative

During the adoption of the Digital Single Markets strategy on 6 May 2015, the European Commission “announced the launch of a cloud for research” (European Commission, 2016a). The Commission appointed a High Level Expert Group on the European Open Science Cloud to advise on the scientific services that should be delivered in the cloud, and what its governance structure should look like. This group published its first report ‘Realising the European Open Science Cloud’ in 2016 (Commission High Level Expert Group on the European Open Science Cloud, 2016; European Commission, 2016a; European Parliament, 2016). This was followed by a ‘Report Towards a Digital Single Market Act’, which was adopted on 19 January 2016 and dealt directly with the European Open Science Cloud (European Parliament, 2016). The European Cloud Initiative was subsequently initiated in 2016 to “strengthen Europe’s position in data-driven innovation, improve competitiveness and cohesion, and help create a Digital single market in Europe” (European Commission, 2016b). Through this initiative, the EU created and established a European Open Science Cloud that would present Europe’s researchers and science and technology professionals with a “virtual environment” where data can be stored, shared and re-used “across disciplines and borders” (Claudet et al., 2016). This Cloud would, over time, be enlarged and opened up to the public sector and to industry. Underpinning this is the European Data Infrastructure (EUDAT), which is “deploying the high-bandwidth networks, large-scale storage facilities and super-computer capacity” that will be needed for effective accessing and processing of huge datasets in the Cloud (Claudet et al., 2016).

The European Cloud Initiative is being rolled out through a number of actions:

- 2016: Creation of a European Open Science Cloud;
- 2017: “Opening up by default of all scientific data produced by projects” under the Horizon 2020 research and innovation programme;
- 2018: “launching of a flagship initiative to accelerate” the advancement of “quantum technology”; and
- 2020: development and deployment of a wide-ranging “European high-performance computing (HPC) data storage and network infrastructure,” which will include “two prototype next-generation supercomputers,” establishment of a

"European big data centre," and upgrading of the "backbone network for research and innovation (GEANT)" (Claudet et al., 2016).

4.3.3.5 European Data Portal

The European Data Portal was launched in March 2016 and harvests metadata of public sector information available on public data portals throughout European countries (European Data Portal, 2017). This includes open government data such as information collected and produced for, or funded by public bodies, as well as information held by the public sector. Although this portal does not specifically focus on research data per se, information from this portal might be valuable to researchers (European Data Portal, 2017).

4.3.3.6 Open Research Data Pilot

The Horizon 2020 (EU framework programme for research and innovation) affords research projects an opportunity to participate in an Open Research Data Pilot (Johnsson and Åhlfeldt, 2015: 13). The Open Research Data Pilot "aims to improve and maximise access to and re-use of research data generated by projects" (European Commission, 2016c). For 2014-2015, a number of research areas participated in the Open Research Data Pilot. These were:

- "Future and Emerging Technologies;
- Research infrastructures – part e-Infrastructures;
- Leadership in enabling and industrial technologies – Information and Communication Technologies;
- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities;
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials;
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies; and
- Science with and for Society" (OpenAIRE, 2016).

4.3.3.7 Other significant RDM initiatives in the EU

As mentioned in 4.3.3.2, the High Level Expert Group on Scientific Data submitted a report in 2010, called 'Riding the wave: how Europe can gain from the rising tide of scientific data', to the European Commission. The report focused on potential scenarios for future European researchers (High Level Expert Group on Scientific Data, 2010: 13-15; Johnsson and Åhlfeldt, 2015: 13). The report listed a number of objectives and actions that would need to be implemented to ensure the establishment of e-infrastructures for research in Europe (High Level Expert Group on Scientific Data, 2010: 29-33; Johnsson and Åhlfeldt, 2015: 13). It contained clear and comprehensive objectives for member states of the EU, touching on issues such as preservation and re-use of data, e-infrastructures for research data, and open access to research data (High Level Expert Group on Scientific Data, 2010: 23, 34-35; Johnsson and Åhlfeldt, 2015: 13).

Another important initiative in RDM has been the collaboration between universities in Europe, called the League of European Universities (LERU). LERU published the 'LERU Roadmap for Research Data' in December 2013, which concentrated on issues such as policy research data infrastructure, costs, leadership, advocacy, skills, roles, responsibilities, description and legal issues, and provided clear recommendations and instructions on RDM to LERU member universities (Johnsson and Åhlfeldt, 2015: 13; LERU Research Data Working Group, 2013). The Roadmap presented a number of recommendations:

- **For institutional policy and decision makers**: Including formation of RDM Steering Groups at individual LERU members, the development of institutional Roadmaps for Research Data, development and promulgation of institutional data policies, cost modelling for RDM, creation of data management support services, introduction of specific RDM positions with career paths, recognition and fostering of data science as a professional discipline, development of policies for promoting and rewarding those that are generating and sharing data; collaboration between LERU universities in RDM, and the continual infusement of policy at the EU level;

- **For those who are involved in the curation of research data**: Involving the placement of research data into the framework of the Opportunities for Data Exchange (ODE) Data Publication Pyramid (showing the categories of data that can be made available for sharing and re-use) in order "to support work on description and curation of data," identification of documentation and metadata requirements from the start of any project, the importance of collaboration between researchers and institutional support staff, the compliance of metadata with existing standards, and the offering of a general framework for research data infrastructure;

- **For researchers and their institutions**, entailing further work to reach consensus in the LERU community, working together in clarifying "what is expected of researchers when citing data," identifying the owner(s) of data, stating the "terms of re-use of datasets", organising practical support to researchers, embedding credited RDM "courses within postgraduate training", "engaging in information activities and data audits" that will raise awareness among researchers and the wide community, involving a wide array of stakeholders in training and development, incorporating "data curation into library school" curricula, investing in quality "continuing professional development", and taking note of DMP requirements by funders;

- **For LERU members and the LERU Chief Information Officers (CIOs)**, including sharing of information between LERU universities on the usage of tools and information identified in the Roadmap, collaboration between universities for shared services, collections and curation to curtail costs, provision of "general information and guidance" on the issue of "open research data", establishment of "doctoral schools for advanced data management" / data science, engagement at international level, fostering of "debate amongst stakeholders and disciplines" on data sharing, developing and clearly articulating "incentives for researchers" to share their data, promoting best practice in RDM, developing a "portfolio of tools" for an institutional infrastructure, establishment of an institutional asset register, and organisation of an "institutional research data workforce"; and

- **For the bodies of the EU**, entailing the encouragement and support by the EU of national stakeholders" to develop RDM policies", the engagement by the EU with universities by facilitating "pan-European approaches" to the issue of RDM "in the context of the European Research Area", "introducing funding opportunities for European Universities" in the area of data-driven science, bridging the gaps in skills development through programmes such as Horizon 2020, and revising the "EU Copyright and Database Directives" to "enable secure Text and Data Mining" (LERU Research Data Working Group, 2013: 13, 31-33).

Yet another initiative was SIM4RDM, a two-year project funded under the EC's Seventh Framework Programme (FP7), which was also launched in 2012 to enable researchers to take full advantage of emerging data infrastructures in the European Research Area by ensuring that they have the knowledge, skills and support infrastructures necessary to adopt good RDM (Mossink, Bijsterbos and Nortier, 2013: 5).

Still another initiative was a survey of directors of its member libraries conducted by the Association of European Research Libraries (LIBER), in collaboration with DataONE (Data Observation Network for Earth) in 2016, to investigate the types of research data services (RDS) that were being offered by European academic libraries, and the services planned for the future (Tenopir et al., 2017: 25-26). Results from this survey showed that European academic libraries are rendering "more consultative / reference type services", for example assisting clients to find information on DMPs, "metadata and data standards, rather than technical" RDM services, for example "identifying data for inclusion" into a repository (Tenopir et al., 2017: 37). The majority of libraries also revealed that they are offering consultative-type RDM services, which included discussions on RDM and planning or developing RDM policies (Tenopir et al., 2017: 37). Currently, less than half have RDM policies in place. In addition, very few libraries indicated that they are rendering technical RDM services (e.g. creating / transforming metadata for data or datasets, preparing data for deposit, de-accessioning data, etc.) or are planning to do so in the future (Tenopir et al., 2017: 37). Furthermore, European libraries were shown to support RDM for a wide array of data types from qualitative to quantitative research. These libraries also reported differences in their levels of engagement in delivering RDM services to staff and students from diverse disciplines (Tenopir et al., 2017: 38). RDM require library staff that are knowledgeable in RDM and the results from the survey disclosed that many of the European libraries are affording staff with opportunities to learn new skills in RDM, whilst others are appointing new staff for RDM (Tenopir et al., 2017: 38). The study also revealed that there is collaboration between these libraries with internal and external partners, in order to address the issue of RDM at their institutions (Tenopir, 2017: 38).

4.3.4 USA

One of the earliest recommendations for open access to data in the USA came in 1997, when the US National Research Council recommended that "full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from research." It also stated that this should be "balanced against legitimate concerns for the protection of national security," privacy of individuals and "intellectual property" (National Research Council, 1997: 10).

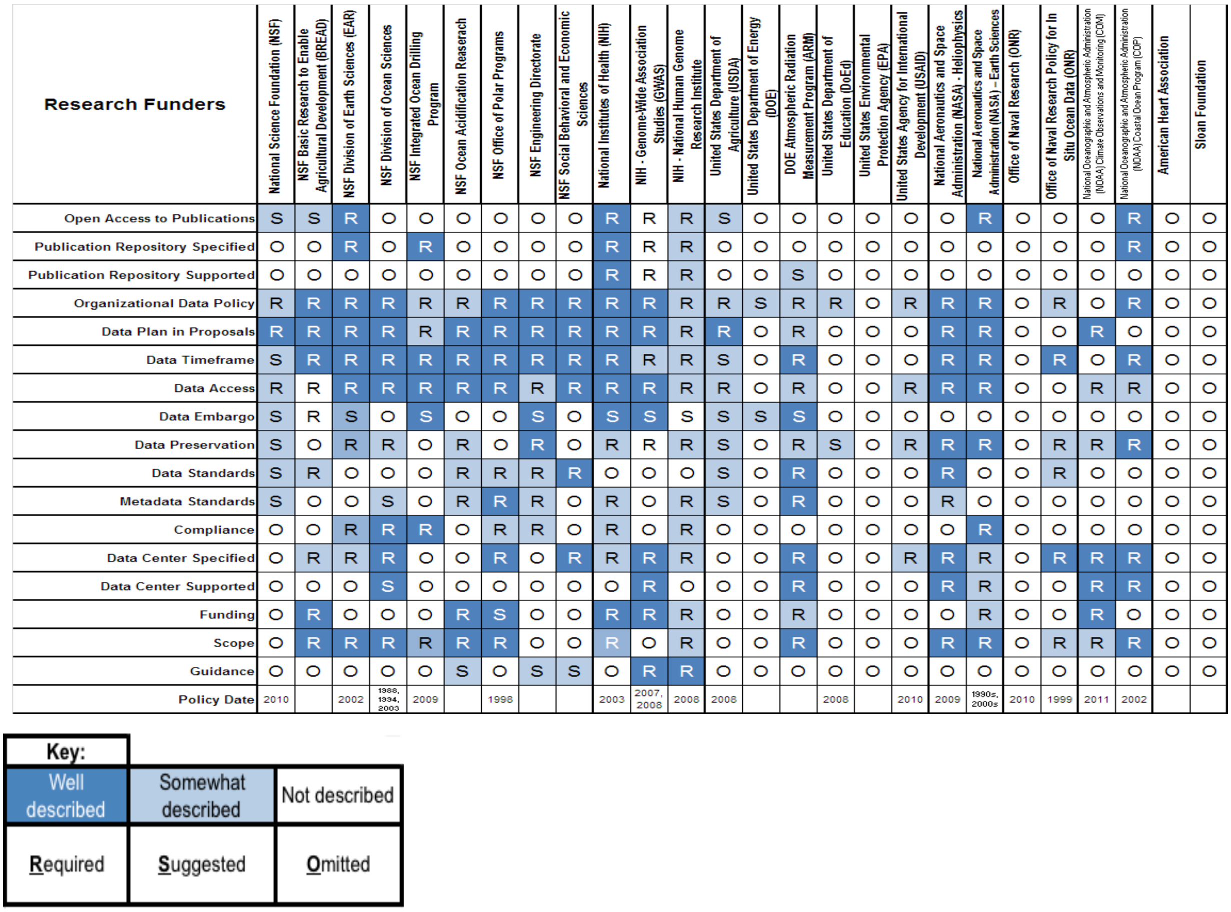
In 2000, the US Congress raised questions abound the inefficiency and duplication of research projects funded by federal agencies in the USA. This led to an investigation by the NSF, resulting in what became known as the Atkins Report (Johnsson and Åhlfeldt, 2015: 12-13). This report focused on how research was being and could be conducted through cyberinfrastructure (see 2.2.2). The report emphasized the importance of data repositories where research data could be preserved and accessed by other researchers across the world (Atkins, 2003: 77). In 2006, the Association of Research Libraries (ARL) published a follow-up to the Atkins Report, with a report titled: 'To stand the test of time: long-term stewardship of digital data sets in science and engineering' (Friedlander and Alder, 2006). This report was the first to give a comprehensive description of the proposed role for research libraries in RDM (Friedlander and Alder, 2006: 11). This was followed by several calls for research projects focusing on infrastructure for data management. Examples of such projects were the DataOne and the Data Curation Profiles Toolkit Projects (discussed below) (Johnsson and Åhfeldt, 2015).

From 2007 until May 2009, the NSF issued the DataNet (Sustainable Digital Data Preservation and Access Network Partners) project. This programme recognised that "science and engineering research and education are increasingly digital and data intensive" (NSF, 2007). It also recognised that "digital data are not only the output of research" but yields "input to new hypotheses," enables "new scientific insights" and drives innovation (NSF, 2007). The key challenge that was identified through this programme was how to develop new methodology, management structures, as well as technologies that would assist in managing the variety, size, and complexity of the existing and "future data sets and data streams" (NSF, 2007). The purpose of the programme was to develop a number of national and global data research infrastructure organisations that could present unique opportunities to communities of researchers to advance science and/or engineering research and learning (NSF, 2007). It was also foreseen that these new organisations would integrate cyberinfrastructure, computer and information sciences, library and archival sciences, and domain science expertise. Programmes that were awarded grants through DataNet were: DataONE, focusing on environmental science; the Data Conservancy, focusing on astronomy, earth science, life sciences and social science; SEAD (Sustainable Environment through Actionable Data), offering data tools to researchers so that they can "easily manage, interpret, share, and publish scientific data to institutional partner repositories"; the DataNet Federation Consortium, focusing on assembling "national data infrastructure that enables collaborative research, through federation of existing data management infrastructure"; and Terra Populus (now known as IPUMS Terra), a global population / environment data network, focusing on tools that could be used for integration, analysis, and visualisation of a wide array of data about "human population" attributes, "land use, land cover, climate and other environmental" features, "that have spatial and temporal dimensions" (DataConservancy, n.d; DataOne: Data Observation Network for Earth, n.d.; DataNet Federation Consortium, 2017; IPUMS Terra, 2016; NSF, 2009; NSF, 2014; SEAD, n.d.). All of these programmes developed into fully fledged research infrastructure organisations that were still providing a national and global service to the research community at the time of this study (DataConservancy, n.d; DataOne: Data Observation Network for Earth, n.d.; DataNet Federation Consortium, 2017; IPUMS Terra; SEAD, n.d.).

Since 2011, the NSF has made a DMP mandatory when submitting grant proposals (Mossink, Bijsterbos and Nortier, 2013: 12). The DataRes project (2011-2012) however found that although most institutions have RDM plans in place to conform to the grant requirements of the NSF, the majority devoted an almost insignificant amount of their budgets to RDM functions, which meant that RDM programs seemed to be mostly conceptual and prospective (Halbert, 2013: 1). There are, however, areas where RDM has been given higher attention, for example the Data Curation Profiles Toolkit (mentioned above), which is the result of a collaboration between the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois Urbana-Champaign, with support from the Institute for Museum and Library Services (Data Curation Profiles Toolkit, n.d.). Another example is the development of the Purdue University Research Repository (PURR), which is a research collaboration and data management solution for Purdue researchers and their co-researchers (Purdue University, 2013). Yet another example is DataONE (Data Observation Network for Earth) (mentioned above), an international collaborative network, supported by the NSF, which was set up to "ensure the preservation, access, use and re-use of multi-scale, multi-discipline, and multi-national science data via three primary cyberinfrastucture elements and a broad education and outreach program" (DataOne: Data Observation Network for Earth, n.d.). The programme focuses on "environmental science through a distributed framework and sustainable cyberinfrastructure meeting the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data" (DataOne: Data Observation Network for Earth, n.d.). A number of institutions in the USA and the UK also collaborated in developing the Data Management Tool (DMP) (California Digital Library, 2014). These are: the University of California Curation Center (UC3) at the California Digital Library, DataONE, Smithsonian Institution, Library of the University of California at Los Angeles, University of California, San Diego Libraries, University of Illinois at Urbana-Champaign, University of Virginia Library, as well as the Digital Curation Centre (UK) (California Digital Library, 2014). The tool is hosted by the University of California Curation Center of the California Digital Library, and is available to researchers to create ready-to-use DMPs for specific funding agencies; to meet requirements for DMPs; to get step-by-step instructions and guidance for DMPs; and to learn about resources and services available at one's institution to fulfil the data management requirements of their grants (California Digital Library, 2014). As mentioned in 4.3.2.3, the DCC and the CDL joined forces in writing a proposal for the building of a new global data management advisory platform, Roadmap, that would include the work done by the DCC in DMPonline and the work done by CDL on the DMPTool.

Dietrich et al. (2012) conducted valuable research in comparing the requirements of funders with regards to RDM. These are tabled in Figure 4.3.

Figure 4.3: Funder Requirements In The USA (Dietrich et al., 2012)



Research Funders	National Science Foundation (NSF)	NSF Basic Research to Enable Agricultural Development (BREAD)	NSF Division of Earth Sciences (EAR)	NSF Division of Ocean Sciences	NSF Integrated Ocean Drilling Program	NSF Ocean Acidification Reaserach	NSF Office of Polar Programs	NSF Engineering Directorate	NSF Social Behavioral and Economic Sciences	National Institutes of Health (NIH)	NIH - Genome-Wide Association Studies (GWAS)	NIH - National Human Genome Research Institute	United States Department of Agriculture (USDA)	United States Department of Energy (DOE)	DOE Atmospheric Radiation Measurement Program (ARM)	United States Department of Education (DoEd)	United States Environmental Protection Agency (EPA)	United States Agency for International Development (USAID)	National Aeronautics and Space Administration (NASA) - Heliophysics	National Aeronautics and Space Administration (NASA) – Earth Sciences	Office of Naval Research (ONR)	Office of Naval Research Policy for In Situ Ocean Data (ONR)	National Oceanographic and Atmospheric Administration (NOAA) Climate Observations and Monitoring (COM)	National Oceanographic and Atmospheric Administration (NOAA) Coastal Ocean Program (COP)	American Heart Association	Sloan Foundation
Open Access to Publications	S	S	R	O	O	O	O	O	O	R	R	R	S	O	O	O	O	O	O	R	O	O	O	R	O	O
Publication Repository Specified	O	O	R	O	R	O	O	O	O	R	R	R	O	O	O	O	O	O	O	O	O	O	O	R	O	O
Publication Repository Supported	O	O	O	O	O	O	O	O	O	R	R	R	O	O	S	O	O	O	O	O	O	O	O	O	O	O
Organizational Data Policy	R	R	R	R	R	R	R	R	R	R	R	R	R	S	R	R	O	R	R	R	O	R	O	R	O	O
Data Plan in Proposals	R	R	R	R	R	R	R	R	R	R	R	R	R	O	R	O	O	O	R	R	O	O	R	O	O	O
Data Timeframe	S	R	R	R	R	R	R	R	R	R	R	R	S	O	R	O	O	O	R	R	O	R	O	R	O	O
Data Access	R	R	R	R	R	R	R	R	R	R	R	R	R	O	R	O	O	R	R	R	O	O	R	R	O	O
Data Embargo	S	R	S	O	S	O	O	S	O	S	S	S	S	S	S	O	O	O	O	O	O	O	O	O	O	O
Data Preservation	S	O	R	R	O	R	O	R	O	R	R	R	S	O	R	S	O	R	R	R	O	R	R	R	O	O
Data Standards	S	R	O	O	O	R	R	R	R	O	O	O	S	O	R	O	O	O	R	O	O	R	O	O	O	O
Metadata Standards	S	O	O	S	O	R	R	R	O	R	O	R	S	O	R	O	O	O	R	O	O	O	O	O	O	O
Compliance	O	O	R	R	R	O	R	R	O	R	O	R	O	O	O	O	O	O	O	R	O	O	O	O	O	O
Data Center Specified	O	R	R	R	O	O	R	O	R	R	R	R	O	O	R	O	O	R	R	R	O	R	R	R	O	O
Data Center Supported	O	O	O	S	O	O	O	O	O	O	R	O	O	O	R	O	O	O	R	R	O	O	R	R	O	O
Funding	O	R	O	O	O	R	S	O	O	R	R	R	O	O	R	O	O	O	O	R	O	O	R	O	O	O
Scope	O	R	R	R	R	R	R	O	O	R	O	R	O	O	R	O	O	O	R	R	O	R	R	R	O	O
Guidance	O	O	O	O	O	S	O	S	S	O	R	R	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Policy Date	2010		2002	1988, 1994, 2003	2009		1998			2003	2007, 2008	2008	2008			2008		2010	2009	1990s, 2000s	2010	1999	2011	2002		

Key:		
Well described	Somewhat described	Not described
Required	Suggested	Omitted

The USA funder requirements model (Figure 4.3) lists a number of requirements that are not included in the UK funder requirements model (Figure 4.2): ‘publication repository specified’, ‘organizational data policy’, ‘data standards’, ‘metadata standards’, ‘compliance’, ‘data embargo’, ‘scope’, and ‘policy date’. The UK Model also lists a number of requirements that are not listed in the USA Model, such as the ‘data’ and the ‘published output’ that the policy covers, and ‘monitoring’ as a policy stipulation. There is, nevertheless, a similarity between a number of the requirements in both models, although some might be worded differently, for example ‘open access’ versus ‘access/sharing’, ‘publication repository’ versus ‘repository’, ‘data timeframe’ versus ‘time limits, ‘data access’ versus ‘access/sharing’, ‘data plan in proposals’ versus ‘data plan’, ‘data preservation’ versus ‘long-term curation’, and ‘funding’ versus ‘costs’. In the UK Model (Figure 4.2), the requirements are divided into three sections: requirements that the funder's policy cover; requirements that the policy stipulates; and the support that is provided by the funder. In the USA Model, however, the requirements are not categorised in any manner. In the UK Model, the key is divided into those requirements that are covered in full, those that are partially covered, and those that are not covered at all. The USA Model, on the other hand, have a dual key - one that indicates with colour codes how a requirement is described (well-described, somewhat described and not described), and another that indicates whether something is required, suggested or omitted.

In 2013, the USA White House Office of Science and Technology Policy (2013: 1), touched on the issue of open access to data by releasing a requirement that "the direct results of federally funded scientific research," which included access to data, should be made available for use by "the public, industry and the scientific community." In the same year, the Association of Research Libraries (ARL) also published an assessment they had done on RDM services in US libraries (Fearon et al., 2013: 11). This assessment revealed that at the time, North American research libraries were "expanding or adopting new research data services," but these were only "in the early stages of development and implementation" (Fearon et al., 2013: 11). In 2014, the Digital Library Federation (DLF) launched the DLF eResearch Network (eRN) as a community of practice that focuses on implementing RDM services and developing RDM skills and collaborative capacity among its members (DLF, n.d.). The purpose was to assist staff from academic and research libraries in developing strategies to create and implement e-Research, digital scholarship, and RDM support services "through a peer-driven, shared learning experience and collaborative projects" (DLF, n.d.).

Most of the major federal grant funders in the US (including the NSF, the National Institutes of Health (NIH) and the Department of Energy) had by 2015 implemented data management and sharing policies; however there seemed to be a lack of common standards for RDM and archiving, as well as a lack of common requirements and enforcement practices for sharing data across agencies (Flores et al., 2015: 86). A number of cross-institutional partnerships, nevertheless, were shown to have been developed through the e-Science Institute (sponsored by the ARL, DLF and DuraSpace); the DLF eResearch Network; the Association of College and Research Libraries (ACRL); Data Management Working Group; the New England Collaborative

Data Management Curriculum; and the Virginia Data Management Bootcamp (Flores et al., 2015: 89).

On 8 February 2016, the United States Geological Survey (USGS), a contributor to the DataONE project, published a press release where they stated their “commitment to open data” as part of their new public access plan (DataONE and USGS: making open data a reality, 2016: 1). In this plan, they pledged to “expand their current on-line gateways” to render “free public access to scholarly research and supporting data produced in full or in part with USGS funding” (DataONE and USGS: making open data a reality, 2016: 1). Exceptions were only allowed in cases of security, privacy and confidentiality (DataONE and USGS: making open data a reality, 2016: 1). This plan corresponded to the statement in the White House Office of Science and Technology Policy memorandum mentioned earlier.

A more recent development has been the establishment of the National Data Service (http://www.nationaldataservice.org), which is a consortium that was formed to create an “open framework that supports an integrated set of national-scale services that will, individually and collectively, enable the efficient, convenient, and secure storage, sharing, publication, discovery, verification, and attribution of data by individuals, groups, and large collaborations” (The National Data Service: a vision for accelerating discovery through data sharing, 2014). This consortium links NSF DataNet (Data Conservancy, DataONE, SEAD), DIBBs (NCSA Brown Dog) and other major disciplinary initiatives (e.g. ICPSR, ADS); MREFCs (IceCube, LIGO, LSST, NEON), universities, and national organisations and services that connect them (Globus, Internet2, XSEDE, SHARE); publishers (e.g., APS, Elsevier, Nature, Science); and important international efforts (e.g., RDA, Helmholtz, EUDAT, OpenAire). The aims of the National Data Service are:

- “To develop a set of common services that can build on top of existing standards and infrastructure provided by various communities;
- To help researchers find data, which entails: cross-disciplinary searching across federations, projects, archives, and other repositories; and finding data related to a publication; as well as drilling down to leverage specialized community-specific discovery;

- To help researchers use data, which includes downloading of data, browsing for metadata, tracking of provenance, as well as moving data to processing platforms for specialized (re-)processing and analysis;
- To help researchers share and publish data, which comprise engaging researchers early in the publishing process; the development of a NDS Repository as a platform for publication preparation; private sharing with collaborators prior to publishing; tools to help organize the data for publishing; automatically ensuring links to literature; assignment of Digital Object Identifiers (DOIs) to the data; provision of links to publishers; synchronisation of data publishing with papers; and recommending of appropriate discipline/community repositories for long-term preservation; and
- To help researchers make data citable" (The National Data Service: a vision for accelerating discovery through data sharing, 2014).

4.3.5 Australia

Unlike the American approach, the Australian RDM programme has not been driven by mandates placed by funding organisations. The Australian government invested a huge amount of funding in research infrastructure and research data initiatives through for example the Australian National Collaborative Research Infrastructure Strategy (NCRIS) and the Australian National Data Service (ANDS) (Treloar, 2009: 126). Investment in research infrastructure by the Australian government started in 2004 in response to the release of the final report of the National Research Infrastructure Taskforce in 2004, and the issuing of the first National Collaborative Research Infrastructure Strategy (NCRIS) Strategic Roadmap in 2006 (Australian Government, Department of Education, Science and Training, 2004; Australian Government, Department of Education, Science and Training, 2006).

4.3.5.1 National Collaborative Research Infrastructure Strategy (NCRIS)

The Australian Government set the following objectives for NCRIS:

- Provide major research infrastructure that is national, collaborative and world class;

- Promote a sustained cultural shift towards investment attitudes that are national, strategic and collaborative;
- Foster research activity that are world-class; and
- Improve and/or establish agreed priorities for national research infrastructure (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 5; Searle et al. 2015: 441).

NCRIS received funding in the financial year of 2004/5 and its programmes started in the financial year 2006/7. The NCRIS plan included fifteen areas of identified research capability, namely: evolving bio-molecular platforms and informatics; integrated biological systems; characterisation; fabrication; biotechnology products; translating health discovery to clinical application; population health and clinical data linkage; Networked Biosecurity Framework; heavy ion accelerators; optical and radio astronomy; Terrestrial Ecosystem Research Network; Integrated Marine Observing System; structure and evolution of the Australian continent; low-emission, large-scale energy processes; next generation solutions to counter crime and terrorism; and platforms for collaboration (Australian Government, Department of Education, Science and Training, 2006). The platforms for collaboration constituted a number of distinct services:

- Addressing the need for continuous network investment through the Australian Research and Education Network (AREN);
- “A coordinated approach to authentication” through the Australian Access Federation (AAF);
- “Collaboration and middleware” through the Australian Research Collaboration Services (ARCS);
- “HPC” through the National Computational Infrastructure (NCI); and
- “Data management and sharing” through ANDS (Treloar, 2009: 126).

The NRCIS Strategic Roadmap was reviewed and re-issued in 2008. This new Roadmap indicated that the capabilities identified in the 2006 Roadmap would continue to be priorities, but that a reshaping might be needed to accommodate a number of additional needs, and to supplement elements in specific capabilities (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 9). The report further acknowledged the importance of Humanities, Arts and Social Science as an important capability area, as well as the significance of information and communication technologies as an undergirding and prevalent capability (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 9). In addition, the report highlighted the inclusion of data as collaborative infrastructure (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 9).

The 2008 report, furthermore, intentionally grouped the priority research capability areas with the aim to further drive cross capability collaboration. In some instances, clear linkages between capabilities already existed, for example those capabilities with a health or environmental context (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 17). In other instances, sub-components of capabilities were "shifted between capabilities to reflect intuitive or existing" relationships (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 17). This report also mentioned that work was underway for the development of the Australian Research Collaboration Services (ARCS) and the Australian National Data Service (ANDS), which would form a foundation for dealing with "future infrastructure and practices needed to collaborate", and specifically with regards to the sharing, re-using and curation of data (Australian Government, Department of Innovation, Industry, Science and Research, 2008: 17). "The Australian government reaffirmed its commitment to national infrastructure through the National Innovation and Science Agenda (NISA)" in December 2015 (Australian Government, Department of Education and Training, 2016: 2). This allowed for the release of funding for the "NCRIS funded facilities and projects" that existed at the time, as well as the Australian Synchrotron and the Square Kilometre Array (SKA) (Australian Government, Department of Education and Training, 2016: 2).

Currently the NCRIS network facilitates national research capability through 27 active projects, and NCRIS facilities are used by over 35,000 researchers nationally and internationally (Australian Government, Department of Education and Training, 2017). These projects are: Astronomy Australia Ltd; Atlas of Living Australia (ALA); AuScope; Australian Health Laboratory (AAHL); Australian Microscopy and Microanalysis Research Facility (AMMRF); Australian National Data Service (ANDS); Australian Phenomics Network (APN), Australian Plant Phenomics Facility; Australian Plasma Fusion Research Facility (APFRF); Australian Urban Research Infrastructure Network

(AURIN); Biofuels; Bioplatforms Australia (BPA); European Molecular Biology Laboratory (EMBL) Australia; Groundwater; Heavy Ion Accelerators (HIA); Integrated Marine Observing System (IMOS); National Computational Infrastructure (NCI); National Deuteration Facility (NDF); National eResearch Collaboration, Tools and Resources (NeCTAR); National Imaging Facility (NIF); Nuclear Science Facilities (NSF) – Bragg Institute; Pawsey Supercomputing Centre (Pawsey); Population Health Research Network (PHRN); Research Data Storage Infrastructure (RDSI); Terrestrial Ecosystem Research Network (TERN); and Translating Health Discovery (THD) (Australian Government, Department of Education and Training, 2015).

4.3.5.2 Australian National Data Service (ANDS)

Searle et al. (2015: 442) suggest that the Australian Partnership for Sustainable Repositories (APSR) was an important pre-cursor to ANDS. APSR funded a wide array of projects in partnership with “research communities, information professionals, technical staff, and higher education policy makers,” with the purpose to assist in the creation of systems “required for managing data and information” in the Australian research environment, at the same time simultaneously increasing “the capability of Australian researchers to do so” (Australian Partnership for Sustainable Repositories, n.d.). In 2007, APSR conducted the first investigation on data management practices at Australian Universities, which led to a wide array of “community building activities,” which included “training events and mailing lists” (Henty et al., 2008: iv; Searle et al., 2015: 442). The “Online Research Collections Australia (ORCA) software, as well as the Registry Interchange Format-Collections and Services (RIF-CS) schema, a profile of ISO 2146,” was also developed by APSR (Searle, et al., 2015: 442). When ANDS was created, ORCA and RIF-CS was transferred to ANDS to form the basis for its further development (Searle, et al., 2015: 442). The creation of ANDS arose from a number of consultation meetings that were held in 2006 and early 2007, where the need to improve data management and availability came up as a consistent theme (Treloar, 2009: 127). In April 2007, an ANDS Technical Working Group (ANDS TWG) was set up. This working group produced a report in October 2007, that presented a vision stipulating how ANDS might operate (Treloar, 2009: 127). Four programmes of activity within ANDS were envisaged, namely “frameworks, utilities, repositories and researcher practice” (Treloar, 2009: 126).

Monash University, based in Melbourne, in collaboration with the Australian National University, based in Canberra, as well as the Commonwealth Scientific and Industrial Research Organisation (CSIRO), were approached in 2007 by the then Australian Commonwealth Department of Education Science and Training (DEST) to take part in a project to establish ANDS (Treloar, 2009: 127). The project started in January 2008 and was concluded in late 2008, with ANDS coming into existence in September 2008. Funding was provided by the Australian government (Treloar, 2009: 127-128). The aim of ANDS is to partner with "research and data producing" institutions such as Australian universities to manage research data within Australia (Searle et al., 2015: 442). The manner in which ANDS has supported partners, is "through the funding of projects to support the capture, description and storage of data and metadata"; the provision of advice on implementation of software that have been developed "as part of ANDS-funded projects"; and the creation of a community amongst its partners (Searle et al., 2015: 442). This partnering has further led to an acceleration of support for RDM within Australian university libraries (Searle et al., 2015: 442). Searle et al. (2015: 442) further mention that ANDS and the Council of Australian University Libraries (CAUL), have a strong relationship. ANDS also played an important role in promoting institutional RDM policies. In 2011, ANDS released its RDM Framework: Capability Guide, which promoted the development of institutional policies and procedures that address data management requirements (Australian National Data Service, 2011).

ANDS oversaw a number of programmes:

- Public Sector Data (2010-2014), which aimed to make data and related metadata from government departments available;
- Data Capture (2010-2013), which was aimed at "simplifying the process of capturing data and rich metadata" at the point of creation or close to it, and streamlining the "deposit of these data and metadata into well-managed stores";
- Seeding the Commons (2010-2014), which was aimed at improving the environment for data management in a manner that would increase the quantity of content in the data commons (open shared data available for access), and also focused on "embedding skills and services within institutions" to facilitate cultural transformation "among researchers and units that support them";

- National eResearch Architecture Taskforce (NeAT) (2010-2012), which initiated projects "as part of NCRIS under Platforms for Collaboration," which were designed in such a manner that they could "develop infrastructure that responded directly" to the requirements of specific discipline communities;
- Metadata Stores, (2012-2014), which "assisted institutions and disciplines" to more effectively "manage the collection and object level metadata" related to the "research data outputs and associated entities";
- Major Open Data Collections (2014-2015), which developed international meaningful collections that provided institutions with an opportunity to add value to research data assets, enabled the formation of new partnerships, and enabled "new data intensive approaches" that could deal with important research challenges; and
- eResearch Infrastructure Connectivity (2015-2016), which assisted in creating improved connections between capabilities provided by the disparate national eResearch infrastructure providers and the requirements of data intensive NCRIS capabilities (Australian National Data Service, n.d.(c), Norman and Stanton, 2014: 254). This entailed "connections between storage," the development of "data-focused compute services," and descriptions (metadata) made accessible through Research Data Australia (Australian National Data Service, n.d.(c)).

Currently, ANDS is led by Monash University in collaboration with the Australian National University, as well as CSIRO. Monash University was chosen because of its leading role in a number of former e-Research projects, e.g. ARROW (Australian Online Research Repositories to the World), which ended in 2008; DART (Dataset Acquisition, Accessibility, and Annotation e-Research Technologies), which ended in June 2007; and ARCHER (Australian Research Enabling Environment), which ended in 2008. It was also chosen because of its early commitment to institutional data management action (Treloar, Choudhury and Michener. 2012: 174).

ANDS provides a nationally significant resource through which researchers can easily publish, discover, access and use Australian research data. This is done through the following:

- Trusted partnerships: working on research data projects and collaborations with partners and communities;
- Reliable services: rendering national services that "support data discovery, connection, publishing, sharing, use and re-use";
- Enhanced capability: building the data skills and capacity of Australia's researchers (Australian National Data Service, n.d.(a)).

ANDS further provides a web portal called Research Data Australia (available at https://researchdata.ands.org.au), which aids in discovering Australian research data collections from over one hundred Australian research institutions, government agencies and cultural institutions (Australian National Data Service, n.d.(b), Research Data Australia, n.d.). Libraries across Australia "worked with researchers to populate" the Research Data Australia repository (Kingsley, 2016). ANDS, however, does not store the data themselves in this portal, but displays descriptions (metadata) of the data, as well as links to the data held by their partners and contributors (Australian National Data Service (n.d.(b)).

4.3.5.3 Data Storage Infrastructure (RDSI) Project

In the late 2000s, the issue of data storage arose. The Australian government allocated AUS$50 million in its 2009/10 budget for the Research Data Storage Infrastructure (RDSI) Project (RDSI project 2010-2015, 2015). This project started in 2010 and concluded in 2015. The purpose of the RDSI project was:

- To "develop a national network of data stores", called nodes (Paz, 2015);
- To "create and develop a data storage infrastructure" that could be accessible via existing infrastructure, supplied by other agencies working in this sector (Paz, 2015);
- To connect this storage infrastructure to the Australian Research and Education Network (AREN), Australia's advanced research and education telecommunication network, which provides essential, very high-speed and high bandwidth connections between Australian universities and research institutions (Paz, 2015; Tizard, 2014). AREN was funded through the Federal government, built through the National Research Networks (NRN) Project, and launched in November 2014 (Tizard, 2014). In South Australia, AREN was implemented by

SABREnet, and in Victoria, AREN was implemented by VERNet. In other states, nationally and internationally, it was implemented through AARNet (Tizard, 2014). This includes dedicated high-speed connections between eight RSDSI nodes (Paz, 2015; RDSI project 2010-2015, 2015);

- To support exemplary data collections through the Research Data Service Programme (ReDS) (Paz, 2015). The ReDS programme was set up in 2012 to "identify research data holdings of lasting value and importance and contribute funding to their development" at the most appropriate nodes (Research Data Services, 2014); and
- To encourage economies of scale through the Vendor Panel Programme (VePa), a programme that established a vendor panel for use by RDSI and the sector (Paz and Tate, 2014).

4.3.5.4 Other Significant Developments In Australia

In 2007, the Australian Code for the Responsible Conduct of Research was issued by the national grant funding agencies. This Code allocated a "shared responsibility" to "researchers and their institutions" to manage their research data and primary materials well (Searle et al., 2015: 445). It also touched on aspects of storage, ownership, retention and accessibility (Searle et al., 2015: 445). In 2013, the Australian Research Council (ARC) issued an Open Data Policy, and although the policy initially only applied to publication outputs, the ARC's Discovery Projects Funding Rules for 2014-2015 stated that the "ARC considers data management planning an important part of the responsible conduct of research and strongly encouraged the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository" (Australian Government, Australian Research Council, 2014: 18). In 2014, The ARC instituted a new requirement for applications for grants from 2014 onwards, which stipulated that researchers applying for grants under the National Competitive Grants Program should include DMPs in their grant applications (Australian Government, Australian Research Council, 2015; Kennan and Markauskaite, 2015: 70). This was followed by another Australian research funder, the National Health and Medical Research Council (NHMRC), which issued a targeted consultation draft of Principles for Accessing and Using Publicly-funded Data for Health Research in 2014, with the purpose "to maximise the research use of publicly funded health and health-related data" (Kennan and Markauskaite, 2015: 70; Australian Government, National Health and Medical Research Council, 2014).

In 2015, the Australian government released a Public Data Policy Statement which, although not focusing on research data per se, had an impact on the research sector in Australia (Australian Government, 2015). This statement stressed the importance and value of data collected and held by the Australian government (public data), as well as the importance of effectively managing this as a strategic national resource (Australian Government, 2015). In this statement, the government committed to optimise the use and re-use of this data, and to release non-sensitive data as open by default (Australian Government, 2015). The government also committed to collaborate with the research and private sectors to enhance the value of public data, for the benefit of Australian society (Australian Government, 2015).

In 2016, the Australian government released its National Research Infrastructure Roadmap, which identified "priority research infrastructure" for the next ten years in nine focus areas that will support "research in which Australia can and needs to excel" for "long-term national benefit" and the fostering of "strategic international partnerships" (Australian Government, Department of Education and Training, 2016: 1). The nine focus areas are: Digital Data and eResearch Platforms; Platforms for Humanities, Arts and Social Science; Characterisation; Advanced Fabrication and Manufacturing; Advanced Physics and Astronomy; Earth and Environmental Systems; Biosecurity; Complex Biology; and Therapeutic Development (Australian Government, Department of Education and Training, 2016: 3). The Report also recommended the establishment of "a National Research Infrastructure Advisory Group" to render independent advice to the Australian government "on future planning and investment" with regards to research infrastructure. In addition, the Report recommended the development of "a Roadmap Investment Plan" addressing "the needs of complimentary initiatives" such as the "Medical Research Future Fund (MRFF) and the Biomedical Translation Fund (BTF)"; the recognition of the crucial need for a skilled workforce; the recognition of the ongoing requirement for investment in existing Landmark Facilities, for example, "the Australian Animal Health Laboratory (AAHL), the Australian Synchrotron, the OPAL Research Reactor and the Marine National Facility (MNF) RV Investigator"; the implementation of "a coordinated approach to international engagements"; increasing the awareness of national research infrastructure; and addressing national HPC needs as a matter of urgency (Australian Government, Department of Education and Training, 2016: 3-4).

4.3.6 Comparisons Between RDM Developments In The UK, EU, USA And Australia

The earliest development of RDM services occurred in the UK with the launch of the SSRC Databank (Social Science Research Council Databank), a precursor of the UK Data Archive in 1967. The RDM initiatives in the USA have mostly been driven by mandates received from the various funders, for example the NSF, Wellcome Trust, etc. while RDM initiatives in the UK have been driven by funding received from government (through JISC for initiatives such as the UK Data Archive, DCC, and through the ESRC for the UK Data Service), as well as mandates received from funders. RDM initiatives in Australia and the European Union, however, have been mostly driven by funding and development of infrastructure for RDM that were provided by government. In Australia, this was done by providing funding for development and maintenance of research infrastructure through NCRIS and funding for RDM projects in research institutions through ANDS, and in the EU through the Horizon 2020 European Strategy Forum on Research Infrastructures (ESFRI) and the Open Research Data Pilot, as well as the SIM4RDM, a 2-year project funded under the EC's Seventh Framework Programme (FP7). Both the UK and USA, on the other hand, contributed to valuable tools that can assist in the RDM process, for example the DMPonline tool and DMPTool (tools that can help in creating a DMP), as well as the Data Curation Profiles Toolkit (a tool that can help in capturing and organising the data through an interview process). The UK's DCC also have some valuable materials that can assist in the RDM process.

With regards to setting up a data management policy, the UK as well as Australia have very useful examples and materials available to assist researchers and institutions.

The discussion also showed that libraries in the USA (through ARL and DLF), UK (through RLUK), EU (through LiBER) and Australia (through ANDS) have been active in developing RDM services for their institutions. Valuable research has furthermore been done in the UK and the USA to compare the requirements of funders with regards to RDM, as discussed in 4.3.4.

RDM, however, is not only a national issue, but is increasingly a matter of international concern, which has led to the formation of a number of international collaborative initiatives.

4.3.7 International Collaborative Initiatives

4.3.7.1 Committee On Data For Science And Technology (CODATA)

CODATA is an interdisciplinary scientific committee of the International Council for Science (ICSU). It was established by ICSU in 1966, with the aim to improve the quality and accessibility of data, as well as the methods by which data are acquired, managed, analysed and evaluated (CODATA, 2017). This includes every type of data emanating from experimental measurements, observations and calculations in all fields of science and technology (CODATA, 2017). Furthermore, CODATA facilitates "international cooperation" between those collecting, organising and utilising data; promotes "an increased awareness in the scientific and technical community of the importance of these activities"; and deliberates "data access and intellectual property issues" (CODATA, 2017). CODATA achieves its aims and objectives through task groups, working groups, national member activities, standing committees, conferences, workshops, publications, and co-operation with other organisations on collective interests (CODATA, 2017).

4.3.7.2 World Data System

The World Data System (WDS) is an interdisciplinary body of the International Council for Science (ICSU) that was created in 2008, with the following goals:

- Enable universal and equitable access to quality-assured scientific data, data services, products and information;
- Ensure long-term data stewardship;
- Foster compliance to agreed-upon data standards and conventions; and
- Provide mechanisms to facilitate and improve access to data and data products (ICSU, n.d.).

The WDS promotes universal and equitable access to, and long-term stewardship of, quality-assured scientific data and data services, products, and information. The WDS covers the natural- and social sciences as well as the humanities, and coordinates trusted scientific data services for the provision, use, and preservation of relevant datasets (ICSU, n.d.). To fulfil its mandate, the WDS strives to develop worldwide 'communities of excellence' for scientific data services. This is done by certifying member organisations (holders and providers of data or data products) from a wide range of fields using internationally recognised standards (ICSU, n.d.). These members then form "the building blocks of a searchable common infrastructure with which to form a data system that is both interoperable and distributed" (ICSU, n.d.). Membership in WDS (which is sanctioned by ICSU) advances local and international scientific recognition, and also enlarges exposure to potential international users and collaborators. Membership signals a strong and tangible commitment to open data sharing, data and service quality, and preservation, which are attributes increasingly required by funders (ICSU, n.d.). The WDS works closely with CODATA.

4.3.7.3 Research Data Alliance

The Research Data Alliance (RDA) was formed in March 2013, with the aim of building social and technical bridges that would enable the open sharing of data; in other words, the vision is "researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society" (RDA, n.d.). The Research Data Alliance was formed initially by the Australian Government through the Australian National Data Service (ANDS), by the European Commission through the RDA Europe project, and by the USA through the RDA/US, and is still being supported by these countries (RDA, n.d.). "The Research Data Alliance enables data to be shared across barriers through focused Working Groups and Interest Groups, formed of experts from around the world – from academia, industry and government. Participation in RDA is open to anyone who agrees to its guiding principles" (RDA, n.d.).

4.3.7.4 DataCite

DataCite is a global non-profit organisation formed in 2009 in London, with the following aims: to "establish easier access to research data on the Internet"; to "increase acceptance of research data as legitimate, citable contributions to the scholarly record"; and to "support data archiving" that will allow for "results to be verified and re-purposed for future study" (DataCite, n.d.). DataCite, through collaboration, supports researchers by helping them to find, identify and cite research datasets. DataCite also supports data centres by issuing persistent identifiers for datasets, workflows and standards for publication, and further provides support to journal publishers by making it possible to link research data articles to the underlying data (DataCite, n.d.).

4.3.7.5 International Federation Of Data Organisations For Social Science (IFDO)

IFDO was established in 1977 in response to the research needs of the international social science community, with the aim to co-ordinate world-wide data services and thus enhance social science research. IFDO retains associate membership in the International Social Science Council of UNESCO. IFDO's objectives are to:

- "Promote and work for open access to digital data;
- Advocate the preservation of valuable digital resources;
- Support the development of standards, procedures and tools enhancing data usage; and
- Promote and support the establishment and development of data organisations to further these objectives (IFDO, n.d.).

Over the years, IFDO has been publishing a number of survey reports and guides dealing with data archiving, data sharing, data policies and data preservation (International Federation of Data Organisations for Social Science, n.d.). The last report was published in 2014. It was based on a survey that rendered an "overview of data management trends, data policies and data sharing practices" worldwide, specifically in the social sciences (Kvalheim and Kvamme, 2014: 4). The report discussed trends in individual countries, and pointed out some of the challenges some of these countries were facing, especially in the areas of "policy enforcement and data sharing in practice" (Kvalheim and Kvamme, 2014: 4). The report showed that about 50% of the countries surveyed were confronted with researcher funders' and other research stakeholders' data sharing requirements, but did not regularly receive assistance or motivation to fulfil these requirements (Kvalheim and Kvamme, 2014: 23). Data repositories could however be found in many of these countries, which indicated that data sharing was happening (Kvalheim and Kvamme, 2014: 23). In 2016, IFDO sent out a survey to get input from its members with regards to its future direction and new service models for IFDO.

The overview of each of these international collaborative initiatives revealed that some of them had been in place for a long time, for example CODATA since 1966, and IFDO since 1977. The impact of these institutions nonetheless only started accelerating with the development of the Internet and the accompanying digital revolution. The discussion disclosed that CODATA continues to play an important role in fostering international cooperation and collaboration with regards to RDM, but also has an impact in individual countries through national member activities, task groups, workshops and conferences. The discussion also showed that WDS creates an international network of holders and providers of data or data products that ensures worldwide and equitable access to quality-assured research data, data services, data products, and preservation of data. Open sharing through working groups and interest groups was revealed to be the remit of the RDA, while DataCite was shown to focus on the promotion of data citation standards and provision of persistent identifiers (e.g. DOIs). IFDO was revealed to have played an important role in the social sciences through survey reports and guides dealing with data archiving, data sharing, data policies and data preservation. It currently seems to be in a transition phase.

Some of these international collaborative initiatives have had an impact on a number of local initiatives in South Africa; these impacts are indicated in the discussion below.

4.4 RDM DEVELOPMENTS IN SOUTH AFRICA

4.4.1 The Context

In South Africa, there is an increasing awareness of the importance of RDM in various sectors. In order to discuss RDM developments in South Africa, however, it is important to provide a brief overview of the research and scientific landscape in the country.

The Department of Science and Technology (DST) in South Africa takes the overarching responsibility for scientific research in the country and oversees the management of the country's relatively well-developed science system. South Africa has 26 well established universities and two emerging universities in Mpumalanga and Northern Cape provinces (South Africa Yearbook 2015/16: 149). A number of these universities are currently involved in various stages of developing RDM initiatives. These are: Cape Peninsula University of Technology (CPUT), Nelson Mandela University (NMU), North-West University (NWU); Sol Plaatje University (SPU), Stellenbosch University (SU), University of Cape Town (UCT), University of KwaZulu-Natal (UKZN) University of Pretoria (UP), University of South Africa (UNISA), University of the Western Cape (UWC), and University of the Witwatersrand (WITS) (Bezuidenhout, 2016; University of Cape Town Libraries, 2017; Kallenborn, 2013; Klapwijk, 2014; Lötter, 2014b; Macanda and Rammutloa and Bezuidenhout, 2014; Nelson Mandela University, 2017; Woolfrey, 2014; Mias, 2016; Roos, Mias and Van Rooyen, 2017; Sol Plaatje University Annual Report 2015: 2).

Government entities that have, or might be involved in RDM, or have an impact on RDM, are: Academy of Science of South Africa (ASSAf), Department of Higher Education and Training (DHET), Department of Science and Technology (DST), the National Intellectual Property Management Office (NIPMO), the National Research Foundation (NRF), South African National Parks (SANPARKS), the South African National Biodiversity Institute (SANBI), and Statistics South Africa (Lötter, 2014b).

There are a number of research councils that are currently involved in RDM initiatives. These are: Agricultural Research Council (ARC), the Council for Scientific and Industrial Research (CSIR) and the Human Sciences Research Council (HSRC) (Lötter, 2014b). It is foreseen that this list might grow in the near future. South Africa also has seven national research facilities (each producing data in various formats and volumes), which are managed by the National Research Foundation (NRF). These include: the Hartebeesthoek Radio Astronomy Observatory (HartRAO), iThemba Laboratory for Accelerator-Based Sciences, National Zoological Gardens (NZG), South African Astronomical Observatory (SAAO), South African Environmental Observation Network (SAEON), and the South African Institute for Aquatic Biodiversity (SAIAB) (South Africa Yearbook 2015/16: 371-372).

Other South African entities that are involved in RDM are the Library and Information Association of South Africa (LIASA) (through workshops, conferences and seminars), the Association of South African University Directors of Information Technology (ASAUDIT), and the Network of Data and Information Curation Communities (NeDICC) (eResearch Africa Conference, 2013; LIASA WCHELIG/HELIG/DCC Workshop on Developing Research Data Management Services, 2014; Lötter, 2014b; and NeDICC, n.d.). South Africa is also a member of the International Council for Science (ICSU) as well as CODATA through the NRF. The National Committee for CODATA consists of six members (SA National Committee for CODATA: overview, n.d.). The country is also an OECD (Organisation for Economic Co-operation and Development) signatory to 'Open Access for Publicly Funded Research Data' through the South African Department of Science and Technology (DST) (Quint, 2004). There are furthermore individual memberships of the RDA by people in various South African institutions (RDA, n.d.).

A number of the higher education institutions and councils mentioned above, have conducted situation analyses (by means of surveys) to determine the need and readiness of their researchers/institutions for RDM. These include the CSIR, UCT, SU, UNISA, Wits, and UP (Lötter, 2014b). The importance of proper data management is also generally accepted by the CSIR, UNISA, UCT, UP, the Human Sciences Research Council (HSRC), and the NRF (Lötter, 2014b).

4.4.2 Government Initiatives

Government initiatives with regards to RDM are driven by the National Integrated Cyber-Infrastructure System (NICIS), which is supported by the DST (National Integrated Cyber-Infrastructure System, 2017). Other government entities that are involved in RDM include the NRF, the HSRC, the CSIR and the South African Data Archive (SADA).

4.4.2.1 National Integrated Cyber-Infrastructure System (NICIS)

NICIS is supported by the DST and is being deployed in different tier levels. The Tier 1 node includes national infrastructure, the Tier 2 node includes regional infrastructure, and the Tier 3 node, institutional infrastructure (Moholola, 2016). The current Tier 1 infrastructure consists of the following:

(a) Centre For High Performance Computing (CHPC)

The CHPC focuses on providing South African researchers with world-class facilities for high-end computing. The CHPC is managed by the Meraka Institute of the CSIR (CHPC, n.d.; CSIR, 2011).

(b) Data Intensive Research Initiative Of South Africa (DIRISA)

DIRISA originated from the national Very Large Data Base (VLDB) initiative which was initiated in 2008 and is supported by the DST. It operates "on the basis of the need for and feasibility of providing a national integrated cyber-infrastructure system consisting of services for high-end computing, a high-bandwidth network and data curation (Peters, 2013; Wright, 2016: 1). DIRISA's vision is to offer a data infrastructure layer providing services and support for the curation, preservation, exchange and interoperability of research data generated within the national cyber infrastructure (Peters, 2013). DIRISA's mission is to establish a virtual network of distributed nodes; to work towards a national policy on data deposit and exchange; establish data deposit mandates; to promote open science and open data; data curation; and human capacity development (Peters, 2013). DIRISA's current aims are:

- The implementation of "a certified Tier 1 (national) trusted repository" platform where South African researchers can deposit data, and also the deployment and maintenance of "data services" and VREs that would enable researchers to use this platform;
- The establishment "of federated Tier 2 (regional) data repositories" that would underpin "thematic data intensive research", as well as capacity-building; and
- Formulation "of national strategic frameworks for data intensive research and data stewardship" (DIRISA, 2017).

(c) South African National Grid (SAGrid)

The South African National Grid is a project to provide a national grid computing infrastructure to support scientific computing and collaboration. This project is owned, managed and operated by a federation of universities, national laboratories, and research groups, and coordinated by the Meraka Institute under the DST's cyber-infrastructure programme (South African National Grid, n.d.).

(d) South African National Research Network (SANREN)

SANReN consists of "a high-speed network dedicated to research traffic and research into research networking and broadband infrastructures" (SANReN, n.d.). The roll-out of SANReN started in 2007. It was "being rolled out in a phased manner to connect up to 204 sites across [South Africa] with research networks hosting over 3,000 research and education organisations from all over the world" (SANReN, n.d.). SANREN is an essential component of the national cyberinfrastructure to enable researchers to access, compute, analyse, share and retrieve huge data sets at high speed across the country.

4.4.2.2 National Research Foundation (NRF)

The NRF is one of the primary funders for research projects in South Africa. One of the NRF's earliest endeavours in RDM was in 2009, when they established a national portal, SADA, for digitised materials. This portal is available at http://sada.nrf.ac.za/ (SADA, n.d.). SADA is discussed in more detail in 4.4.2.5.

In January 2015, the NRF released its 'Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research' in which it indicated that researchers, in addition to the depositing of "final peer-reviewed manuscripts" of articles in an institutional repository, should also deposit "the data supporting the publication in an accredited Open Access repository, with the provision of a Digital Object Identifier for future citation and referencing" (NRF, 2015). This Statement has had a tremendous impetus among research institutions in South Africa, and played a role in generating a number of RDM initiatives in higher education institutions across the country. These initiatives are touched on in the discussion on the various higher education institutions below.

4.4.2.3 Human Sciences Research Council (HSRC)

The HSRC is a South African science organisation, which conducts research focusing on “improving understanding of social conditions” (Lötter and Van Zyl, 2015: 338). Funding for its activities comes from government, but also from contracts and grants. Many of the HSRC’s large-scale research projects includes surveys that are “nationally representative and “cross-sectional, focusing on behavioural, attitudinal and health related issues” (Lötter and Van Zyl, 2015: 338). When the HSRC published the first South African National HIV Prevalence, Behavioural Risks and Mass Media Household Survey in the early 2000s, the release was accompanied by substantial pressure from society to “make the data underpinning its research findings available to a wider audience of potential users” (Lötter and Van Zyl, 2015: 338). Following this, an international review panel recommended in 2003 that the HSRC should give thought to data management as a crucial part of its future role (Lötter and Van Zyl, 2015: 338).

In 2006, a core team of data management advocates, each with a sound background in research, RDM, and systems development, started investigating methods that could assist in managing and preserving data better, and also make them available for future use. This was followed by the presentation and discussion of a framework for HSRC implementation during road shows and workshops in 2007 (Lötter and Van Zyl, 2015: 339). In addition, “a dissemination interface linked to project information on the web” was set up in preparation for the distribution of pilot data by the end of 2007 (Lötter and Van Zyl, 2015: 338). By the end of 2008, a new act, the HSRC Act (No. 17 of 2008) was released by the government, which confirmed the purposes and objectives of the HSRC (Lötter and Van Zyl, 2015: 339). The Act, in section 3(g), inter alia, stipulates that the HSRC should “develop and make publicly available new data sets to underpin research, policy development and public discussion of the key issues of development, and to develop new and improved methodologies for use in their development” (South Africa, 2008). The Act also specifies “an imperative in terms of publishing data sets to underpin research, policy development and public discussion of the key issues of development. Access to research data is promoted to accelerate the development of solutions to address the challenges of society and to enable developing researchers to contribute to the corpus of scientific knowledge through secondary data analysis” (HSRC, 2014). In 2010, a new indicator was added to the institutional performance, namely the number of research-generated data sets that had been preserved and made accessible for secondary use. This was followed in 2011 by a new requirement that researchers at the HSRC had to provide a data preservation and sharing plan together with their research protocols that had been lodged for ethics review (Lötter and Van Zyl, 2015: 340; HSRC, 2014). By 2015, according to Lötter and Van Zyl (2015: 340), the HSRC had a number of institutional practices in position to "support a data management culture":

- Processes to support effective data management, that consisted of:
    - A research management framework that emphasizes data management as a pivotal aspect of research planning;
    - Research contracts, that specifically refers to the generation of data sets and their ensuing ownership and management.
    - Curation systems, processes, and guidance that included:
    - A metadata capturing interface grounded on "the Data Documentation Initiative (DDI) standard";
    - A file repository for distribution and preservation;
    - A dissemination interface connected to the HSRC's website;
- Processes pertaining to acquisition of data, preparation of data and documents, composing metadata, preservation of data and dissemination of data; and
- Guidance on preparing data and data-related documents for curation, including procedures for verification, the anonymisation of data, the description of data, and the publishing of data (Lötter and Van Zyl, 2015: 340).

The data curation process at the HSRC is regulated by a Data Sharing Policy, which declares that the greater part of HSRC data will be made available for sharing within 12-36 months after the official conclusion date of the project concerned (HSRC, 2014). Planning for data preservation takes place once protocols are submitted to the HSRC Research Ethics Committee by completing a Data Preservation and Sharing Plan. Funders very often specify specific requirements for a project. Information on these requirements are requested from researchers as part of a data preservation and sharing plan for each research project. Assistance is then provided to researchers by the data curation staff to assist them in their planning for RDM, so that they can meet the funders' requirements. Legislative and funder requirements are also considered in policies and procedures, as well as research contracts (HSRC, 2014). Research programmes deposit the data, together with metadata describing the data. A data deposit form is used as a template to facilitate this process (HSRC, 2014). The HSRC uses an integrated system for their data, and Oracle for their metadata. The metadata is entered onto Oracle by using a system that was developed in PHP. The files are placed in Knowledge Tree (also PHP based). The data can then be accessed on the web through an open source web application framework called Zikula (PHP-based) with a MySQL database (Lötter, 2014a).

4.4.2.4 Council For Scientific And Industrial Research (CSIR)

The CSIR is a research and development institution that was established through an Act of Parliament in 1945 (CSIR, 2017). The CSIR is responsible for “directed, multidisciplinary research and technological innovation” (CSIR, 2017). The South African Parliament is the CSIR’s shareholder under the Minister of Science and Technology (CSIR, 2017).

The CSIR’s Information Services’ first attempt to get involved in RDM occurred in 2010, when the CSIR launched their Cooperative Geographical Information System (COGIS) pilot project (Van Deventer and Pienaar, 2015: 40). At the start, they discovered that it would be necessary to first furnish researchers with the appropriate infrastructure to conduct data science (Van Deventer and Pienaar, 2015: 40). Then only could researchers be convinced to adhere to RDM guidelines. An infrastructure was subsequently put in place to give access to research output and related data. It further encouraged the usage of geo-information in research, facilitated access to the geospatial data, guaranteed compliance to legislation, contributed to an expansion in the quality of research outputs, and facilitated collaboration (Van Deventer, 2015: 40). This project provided the CSIR Information Services with tremendous insight into the RDM challenges that could arise in one research discipline, even though they had not instigated the project (Van Deventer and Pienaar, 2015: 41). Something else that was learned was the significance of context-giving documentation, for example research contracts, as well as publications following, which used a particular data set (Van Deventer and Pienaar, 2015: 41).

The CSIR appointed a data librarian in 2014, who conducted a comprehensive review of RDM practices at the CSIR, in the same year. The first step was to try and identify and understand existing researcher behaviour with regard to RDM at the CSIR, by means of a survey (Patterton, 2016; Van Deventer and Pienaar, 2015: 41). The results showed that research data held by researchers in the CSIR were generally considered as confidential, differed in size, and were mostly saved as text, spreadsheets, and images. In addition, it was found that some data were generated by proprietary systems and also stored in them (Van Deventer and Pienaar, 2015: 41). These findings were then applied in a complete CARDIO-model evaluation (an RDM readiness tool, developed by the Digital Curation Centre in the UK) in conjunction with the CSIR's ICT Department, giving their ICT department a greater understanding of the challenges related to RDM (Van Deventer and Pienaar, 2015: 41). This was followed up with a second survey in 2015, which was conducted to determine the RDM practices of emerging researchers (Patterton, 2016). The CSIR Information Services was still working on an RDM policy for the CSIR at the time of this study (Patterton, 2016).

4.4.2.5 South African Data Archive (SADA)

SADA (an initiative by the NRF) acts as a broker between a range of data providers (e.g. government departments, statistical agencies, opinion and academic institutions, and market research companies), and the research community. The purpose of the archive is to preserve data for future use, and also to add value to the collections. It preserves and secures datasets and corresponding documentation, and attempts to make it as easily accessible as possible for research and educational purposes. SADA has the following objectives:

- "To acquire and catalogue survey data and related information;
- To preserve such data against technological obsolescence and physical damage;
- To provide originators or depositors of data with necessary information in order to ensure high standards of data documentation;
- To re-disseminate such information for use by other researchers, for re-analysis of data, longitudinal and comparative studies, research training, teaching and policy-making decision purposes;
- To formulate policies for the scope and content of data and data preservation;
- To promote the optimal use of data" (SADA, n.d.).

SADA covers a wide range of areas, such as censuses and household surveys, Omnibus and international studies, demographic and health related studies, substance abuse, crime, income and poverty, inter-group relations, labour and business, education and training, and political perceptions and attitudes. Through its extensive network, SADA can channel data and information stored in its databases to interested researchers worldwide, and through its computerised system, it can also obtain data from outside the country for interested researchers in South Africa (SADA, n.d.). In addition, SADA has membership in the International Federation of Data Organisations (IFDO), the Council of European Social Science Data Archives (CESSDA) (of which it is an associate member), the Inter-University Consortium for Political and Social Research (ICPSR), and the International Association for Social Science Information Service and Technology (IASSIST) (SADA, n.d.).

4.4.2.6 South African National Parks (SANParks)

The South African National Parks (SANParks) was formed in 1926, and is the body responsible for the management of a system of 21 national parks across South Africa. SANParks' major priorities are conservation and management of biodiversity and heritage assets (SANParks, 2016; SANParks, 2017). SANParks' contribution to RDM is in the form of a data repository, called SANParks Data Repository (available at http://dataknp.sanparks.org/sanparks/style/skins/sanparks/), which is the primary source for information and research data sets that are collected across the entire SANParks system (SANParks, n.d.). The SANParks Data Repository is a collaborative effort between SANParks and the National Center for Ecological Analysis and Synthesis (NCEAS) in the USA (SANParks, n.d.). The SANParks Data Repository is based on software that was developed by the Knowledge Network for Biocomplexity (KNB). It contains metadata that are based on the Ecological Metadata Language (EML) and the Federal Geographic Data Committee (FGDC) specification (SANParks, n.d.).

4.4.3 National Collaborative Initiatives

4.4.3.1 African Research Cloud (ARC)

The African Research Cloud (ARC) is under development as a functional prototype that will eventually form the African Data Intensive Research Cloud (Taylor, 2016). It consists of an infrastructure as a service (IaaS) cloud system, which hosts tools that underpin various models of RDM, including data storage, -transfer, and -processing, and a wide range of data intensive research activities (Taylor, 2016.). The idea is to establish a cloud solution for a network of South African, African and non-African researchers. Such a solution would afford researchers in Africa, as well as their co-researchers from across the world, access to a storage space for research data, as well as compute facilities (ARC: African Research Cloud, 2017). The initial deployment of the model has already been established at UCT and UWC (Taylor, 2016). On 27-28 October 2016, the Inter-University Institute for Data Intensive Astronomy (IDIA) (discussed in 4.4.3.7) and UP hosted the first African Research Cloud Workshop in Pretoria (Taylor, 2016.). Representatives from UCT, NWU, SPU, UWC, UP, Wits, SKA SA, CHPC, DST and DIRISA attended the workshop (Taylor, 2016.). The focus of the workshop was to:

- Give a technical overview of the project and to plan a roadmap for technical research and development for proceeding into the next phase;
- Give an overview of “plans for science domain strategic demonstration projects in [the] astronomy and biomedical research” fields;
- Discuss the utilisation of the ARC for innovation in teaching, training and learning;
- Discuss “the ARC support model and sustainability”;
- Determine the “long-term goals and strategic directions” of the ARC; and
- Expand the “ARC development partnership” as well as “the strategic science portfolio” (Taylor, 2016).

4.4.3.2 Inter-University Institute For Data Intensive Astronomy (IDIA)

IDIA was launched in 2015 as a partnership between South African universities and industry, to address the emerging challenge of big data in astronomy. The construction of The Meerkat telescope, a precursor of the SKA (see 4.4.5.2) initiated “the astronomy big data revolution in Africa” (IDIA, 2017). Researchers in astronomy, computer science, HPC, statistics and eResearch technologies have been gathered under the banner of IDIA to develop data science capability and solutions as part of the establishment of the SKA (IDIA, 2017). The SKA will involve large teams of international researchers with accompanying large data volumes and powerful processing and analysis needs. Partners in IDIA consist of UCT, UP, UWC, NWU, and SAP Software Solutions (IDIA, 2017). IDIA plans to address the challenges mentioned, by setting up assorted work-packages that would each focus on clear-cut aspects associated with the following issues: big data, distributed data systems and federated cloud infrastructure (e.g. the African Research Cloud), "computing architectures for the processing of large astronomical data sets, visual analytics of big data and data science research for mining and scientific analysis of astronomy data sets" (IDIA, 2017). IDIA also have networks with parastatal and private companies, which comprise ASTRON, the Hartebeesthoek Radio Astronomy Observatory (HartRAO) (mentioned in 4.4.1), IBM-Dome, the National Radio Astronomy Observatory, SAAO (mentioned in 4.4.1), the Square Kilometre Array South Africa (SKA-SA) (see 4.4.5.2), and the South African Astroinformatics Alliance (see 4.4.3.4) (IDIA, 2017).

Initiatives such as this provide an unprecedented opportunity for universities and researchers across South Africa and the world to unite in the pursuit of scientific discovery.

4.4.3.3 The Digitisation And Digital Data Preservation Centre

The Digitisation and Digital Preservation Centre is a collaborative South African digitisation initiative. The centre is hosted at the NRF and has the following aims:

- Render technical digitisation support and services to institutions that are unable to do it themselves, as well as to those that can only do it partially;
- Provide or arrange experts to provide training and support to individuals and organisations that are planning to start with digitisation and digital preservation;
- Coordinate collaborative initiatives to undertake digitisation and digital data preservation among higher education institutions, NGOs, and other organisations that are ready to collaborate; and
- Facilitate the sharing of knowledge through a DSpace Repository, available at http://digi.nrf.ac.za/dspace (NRF, n.d.).

4.4.3.4 Network Of Data And Information Curation Communities (NeDICC)

NeDICC was established as an outflow of earlier work done by the South African Research Information Services (SARIS) project (Page-Ship et al., 2005). In 2008, the 1st African Digital Curation Conference and Workshop was held in Pretoria. At the Conference it was decided, in principle, to formally establish the Network of Data and Information Curation Communities (Van Deventer and Pienaar, 2015: 36). This decision was subsequently ratified at several events and NeDICC was then formally established in 2010 by using the SARIS partners as the base community (Van Deventer and Pienaar, 2015: 36). The aim of this network is to promote the development and use of research data and information curation standards and practices, within the South African and African scientific research community, to ensure the long-term preservation and accessibility of digital research outputs in support of e-Research (NeDICC, n.d.). NeDICC provides a forum for practitioners and managers involved in digital object management practices, to exchange experience, knowledge and expertise and also express alternative views (NeDICC, n.d.). Activities are aimed at promoting communication and collaboration between members of NeDICC. These consist of meetings, seminars, workshops and conferences, where issues of interest or concern can be addressed, the community can be exposed to new development and trends, the community can have opportunities to engage with a wider audience, as well as being exposed to opportunities to showcase work and initiatives. It also provides a space for the development of knowledge and skills of members and promotes awareness and best practices relating to digital preservation, dissemination and use of research outputs (NeDICC, n.d.).

NeDICC subsequently successfully arranged four African Digital Scholarship and Curation Conferences, held in 2009 (Pretoria), 2010 (Gaberone, Botswana), 2011 (Pretoria) and 2013 (Durban) (Van Deventer and Pienaar, 2009; Van Deventer and Pienaar, 2015: 36). Following the 2013 conference, a decision was taken to join forces with the eResearch Africa Conference in future, by developing specifically an RDM track at the conference (Van Deventer and Pienaar, 2015: 36, 37).

Membership to NeDICC is open to all higher education institutions, research councils, and government entities. At the time of this study, representatives from the following institutions regularly attended and presented workshops online, or on location at the CSIR Pretoria campus: CPUT, NWU, SPU, SU, UCT, UNISA, UP, Vaal University of Technology (VUT), Wits University, UWC, the Agricultural Research Council (ARC), CSIR, HSRC, NRF, ASSAF, and DIRISA (NeDICC, n.d.).

4.4.3.5 African Open Science Platform

In December 2016, the pilot phase of the African Open Science Platform was launched at the Science Forum South Africa (SFSA) (CODATA, n.d.). This Platform is an outflow from the Science International Accord on Open Data in a Big Data World, which was launched at the SFSA in 2015 (Science International, 2015). The African Open Science Platform has been supported by the DST, is funded by the NRF, directed by CODATA, and implemented by ASSAf (CODATA, n.d.). The Platform is an Africa-wide initiative that is envisaged to “promote the development and coordination of data policies, data training and data infrastructure” (CODATA, n.d.).

4.4.3.6 Seminar Hosted By DST, HSRC, And UP On 5 November 2012

This seminar, titled ‘Preserving and Providing Access to South African Social Science and Humanities Research Data,’ was held with the aim of bringing together scholars and practitioners from diverse disciplines with an interest in the management of research data in the social sciences and humanities. An explicit goal was to lay the foundation for a roadmap that would address the preservation and dissemination of relevant research data within the South African context (Preserving and providing access to South African social science and humanities research data: science seminar, 2012: 4).

4.4.3.7 The South African Astroinformatics Alliance (SA^3)

SA^3 is a collaboration between three astronomical facilities: SAAO, HartRAO and the SKA-SA. It was formed in 2013, and is managed by the NRF. SA^3 aims to “facilitate access by the South African astronomical community to multi-wavelength astronomical data, as well as tools for dealing with them; and to ensure that data produced by facilities in South Africa are accessible to the international community (in a manner that does not violate any ownership rights); as well as to develop human capital through schools and workshops that introduce people to data and tools of the virtual observatory" (South African Astroinformatics Alliance, 2014). SA³ has been formed to develop data storage, access, visualisation and analysis tools in a coherent manner, taking into account the rapidly changing scale and complexity of requirements and the environment. SA³ will further link observational data, theoretical models and simulations to add to the understanding of the universe. SA³ is also a member of The International Virtual Observatory Alliance (IVOA) (South African Astroinformatics Alliance, 2014).

4.4.3.8 The South African Biodiversity Information Facility (SABIF)

SABIF is a network comprising key national partners and stakeholders who provide data through the SABIF portal (http://www.sabif.ac.za/), as well as the end users of the data. These partners and stakeholders include museums, herbaria, universities, conservation agencies, government agencies and departments, and NGOs (SABIF, 2014). SABIF falls within the Biodiversity Information Management Directorate at SANBI. SABIF's portal is designed as a distributed system. Data providers retain ownership of their databases, which are distributed through the SABIF portal (SABIF, 2014). The databases physically reside with the data providers, and are maintained by them. These providers determine what and how data can be shared via the SABIF portal. SABIF supplies from its side, generic agreements on data use, sharing and ownership, along with data tools for end-user applications. SABIF is a national node of the Global Biodiversity Information Facility (GBIF) (SABIF, 2014).

4.4.3.9 The Western Cape Data Intensive Research Facility (WCDIRF)

In September 2016, a consortium was formed by institutions in the Western Cape to establish a Western Cape Data Intensive Research Facility (WCDIRF), which was approved by the DST as a Tier 2 facility of NICIS (See 4.4.2.1; Ochieng, 2016; Mohohlola, 2016). The aim with this initiative was to considerably expand data-intensive research capacity, by establishing and operating a data-centric high-performance computing facility for data-intensive research, which will concentrate predominantly on research challenges of astronomy in the context of the deployment of the SKA project, but also on bioinformatics and corresponding clinical research (Ochieng, 2016). UCT leads the consortium. Other members are: CPUT, SU, UWC, SPU, and the SKA project (Ochieng, 2016). The expectation is that this facility will form a “platform for developing innovative approaches to research with big data” (Russ Taylor, the SKA Research Chair at UCT and UWC, as quoted by Moholola, 2016).

4.4.4 Initiatives At South African Higher Education Institutions

4.4.4.1 Cape Peninsula University Of Technology (CPUT)

The CPUT Library is part of an internal division at CPUT, called Knowledge, Information and Technology Services (KITS), which in addition consist of the e-Learning and Educational Technology Services, Management Information Systems, Computer and Telecommunication Services, and the Web Development and Innovation Office (Chiware and Mathe, 2015: 1). KITS takes responsibility for the creation of platforms, systems and processes for the management of research data at the institution (Chiware and Mathe, 2015: 1). The development of RDM at CPUT is part of the CPUT Libraries’ e-strategy plan, and the Library’s digitisation, scholarly communication and open scholarship initiatives through its repository (called Digital Knowledge) are seen as an integral part of the research data services, because it provides the possibility to link datasets to publications (Chiware and Mathe, 2015: 1-2). RDM at CPUT is also driven by CPUT’s institutional strategy under the auspices of the Research, Technology, Innovation and Partnerships (RTIP) division, which outlines the Library’s role in RDM support (Chiware and Mathe, 2015: 2). The first RDM initiatives at CPUT were the development of a policy framework with regards to RDM, as well as the development of an RDM Services Roadmap for CPUT. The approach followed was to form an institutional RDM Working Group with representatives from the Library, Research Office, Faculties, and Information and Communication and Technology, the institutional Quality Management unit, Records and Archives Services, the Centre for Postgraduate Studies, research chairs, and heads of research units and centres (Chiware and Mathe, 2015: 4). Through meetings and workshops, the RDM Working Group provided guidance for the development of a RDM policy, which was used to develop a policy framework for RDM at the institution (Chiware and Mathe, 2015: 4).

The development of RDM services at CPUT was done in the context of an e-Research environment, taking into account infrastructure development; information flow and management; communication with researchers; and the development of tools, which are all aligned to the research lifecycle (Chiware and Mathe, 2015: 4).The CPUT Library also conducted a pilot project with one of their research groups, the Institute of Biomedical and Microbial Biotechnology (IBMB) in the form of a survey, to determine the researchers' requirements with regards to RDM. The results of the survey showed that there was a huge need for structured services and tools for RDM in the institution (Chiware and Mathe, 2015: 5-6). In 2015, the CPUT Library established a special skills-development plan, running over a three-year period, in order to develop the skills of their librarians with regards to RDM. At the same time, the Library started developing new roles specifically aligned to managing and developing e-Research platforms (Chiware and Mathe, 2015: 8).

In order to set up a RDM infrastructure at CPUT, the CPUT Library partnered with the Technische Universität München Library in Germany, in the E-Research Infrastructure and Communication (eRIC) project (Chiware and Mathe, 2015: 4). The eRIC project can be accessed at http://eric-project.org/. Institutions in Germany and CPUT are partnering in this project (eRIC: e-Research Infrastructure and Communication, 2017). Its focus is on the entire e-Research support lifecycle (Kallenborn, 2013). The collaborating institutions are sharing a common open source platform (mediaTUM) and exchanging ideas and information among various working groups. This includes improving the platform so that it addresses specific institutional needs (Kallenborn, 2013). eRIC also provides its partnering institutions with services that support the research lifecycle, such as advice on the various literature search and reference management services available at institutions, provision of project management tools via the eRIC Workbench (its electronic lab journal), collection and processing of data in the eRIC Workbench, development of customised tools such as interfaces for automatic data import and filtering, and development and integration of tools for analysis and visualisation (eRIC: e-Research Infrastructure and Communication, 2017).

4.4.4.2 Nelson Mandela University (NMU)

In March 2015, the DST, NMU, Cisco, SKA South Africa and the CSIR established the Centre for Broadband Communication at NMU, with the purpose of developing cutting edge technologies that will assist in the management and synchronisation of the astronomical volumes of data that will be generated by the SKA (SKA, 2015). The Centre has also been tasked to develop human capacity to ensure that South Africa has the necessary skills to support the MeerKAT and the SKA projects (SKA South Africa, 2015; Nelson Mandela University, 2017).

4.4.4.3 Sol Plaatje University (SPU)

SPU, a new university in the Northern Cape, introduced a BSc in Data Science in 2015, becoming the first institution in Africa to introduce a dedicated undergraduate degree in data science. This new programme has been well received (MacGregor, 2015; Sol Plaatje University Annual Report 2015: 2).

SPU's geographical location in Kimberley, close to the site where the SKA will be deployed, led to a partnership with the SKA to develop the high-level intellectual capacity that the project will need. SKA also secured funding for the students on the programme (Sol Plaatje University Annual Report 2015: 15).

4.4.4.4 Stellenbosch University (SU)

SU's Library conducted a pilot survey in 2014 to determine the Research Data Retention practices of their Faculty of Engineering. The aim was that the outputs of this survey would be used as a business case to be sent to their Vice Principal: Research, to get in-principle support for an institutional RDM survey as well as all aspects (including policy) of RDM (Klapwijk, 2014). The subject librarians were also earmarked to do a survey among their respective academic departments to determine which subject data repositories their researchers were using to deposit their data sets for publications, for example Elsevier. The plan was to compare the policies of these subject repositories with the institutional policy they were compiling (Klapwijk, 2014). At the time of the completion of this thesis, it was not clear what the outcomes of initiatives were.

In 2014, SU was also investigating storage platforms for RDM, specifically looking at cloud-based systems (e.g. Unicloud), to be run locally or in consortia (e.g. iRODS) (Klapwijk, 2014). In early 2017, US advertised a position for Manager: Research Data Services, to lead and facilitate strategic innovation in the Research Data Services of the Library. The duties of this position included the following: be an advocate and support for RDM; manage data collections; enhance data security; and act as liaison between researchers, the Library and other role-players, on RDM services (Hendrikse, 2017).

4.4.4.5 University Of Cape Town (UCT)

Around 2012, the UCT Research Committee (URC) initiated a process of institution-wide data curation planning at UCT. The URC created a task team to assist UCT in establishing an effective RDM policy. The task team comprised UCT's ICT Director and the Director of Libraries, together with the Deputy Deans for Research in each faculty, or their nominees. The RDM Task Team (RDMTT) then co-opted a representative from UCT's Research Contracts and Intellectual Property Services (RCIPS) as well as the Director and Manager of UCT's data service, DataFirst. DataFirst's Manager subsequently led the project. A project plan was drawn up and presented to the URC by the project coordinator in April 2013. The URC proposed that the project plan be submitted to UCT's Senate Executive Committee (SEC) for approval, before proceeding further. The SEC in their deliberations addressed issues such as repository needs for such a data collection project, data protection, as well as data sharing. The huge cost implications suggested by the URC for storage infrastructure as well as data collation and metadata creation, were also discussed. The SEC then decided to provide in-principle support for the project, which was to include a draft policy document and implementation plan (Woolfrey, 2014: 1-2).

In order to establish an effective RDM policy a number of actions were taken:

- A scoping study of RDM policies of funding agencies were undertaken by the RCIPS office, in order to draw attention to common themes and isolate best practice. This information was deemed to be an important input into policy creation, because funders' policies weighed heavily on researchers and could provide an incentive for researchers to support RDM. Findings were presented at a RDM Workshop with stakeholders, held on 24 June 2013. The workshop presentation gave an overview of RDM requirements of key funders of UCT research, and findings showed that all the funders required data deposits and encouraged data sharing (Woolfrey, 2014: 2). All but one "required data management plans from researchers, and included time frames for data to be made available to other researchers" (Woolfrey, 2014: 2-3);

- An RDM library working group was formed in June 2013 after the project manager met with library staff. The library RDM working group then conducted a scoping study of RDM policies of publishers of refereed academic journals. Fifty of the peer-reviewed journals in which UCT academics published during 2012, were reviewed. The findings were included in the 'UCT RDM Policy Project Report,' published in March 2014;
- In July 2013, the project manager conducted an e-mail survey among UCT researchers to determine the RDM awareness, needs and practices at the institution (Woolfrey, 2014: 3). Findings were also presented in the UCT RDM Policy Project Report, published in March 2014; and
- From April 2013 – February 2014, the project manager undertook a scoping study to investigate RDM policymaking at other universities. RDM support websites and policy documents of selected universities were explored to identify common issues that needed to be addressed to create a workable policy document (Woolfrey, 2014: 3).

In October 2013, UCT hosted the first 'eResearch Africa Conference' as member of the ASAUDIT in October 2013 (eResearch Africa 2013 Conference, 2013). RDM was one of the themes that were focused on. During the conference, UCT also announced the launch of an eResearch Centre at the University. On 7 March 2014, Lynn Woolfrey of UCT released the UCT RDM Policy Project Report. This report gave an overview of the RDM roadmap followed by UCT. This was followed, on 24-25 March 2014, by the Library and Information Studies Centre sponsoring an RDM workshop to research support staff at UCT. It was presented by Joy Davidson and Sarah Jones from the Digital Curation Centre in the UK (LIASA HELIG and WCHELIG Workshop Organizing Committee et al., 2014).

The Library and Information Studies Centre at UCT has been offering a MPhil specialising in Digital Curation since 2015. This degree covers, among other things, the principles of digital curation, information architecture and metadata, technology enablers for digital curation, and RDM (New MPhil, Specialisation in Digital Curation, 2014). This Centre also presents a number of short courses in data curation and RDM each year (Matlatse, Pienaar and Van Deventer, 2017). UCT, in addition, has a dedicated research data service called DataFirst, which gives researchers access to survey and administrative microdata (data at unit record level) collected from across South Africa and other African countries (DataFirst, 2017).

The RDM entity in the Library and Information Studies Centre resorts in the Digital Library Services section, which consists of six staff members: the Head: Digital Library Services and two digital curation officers, a digitisation officer, and two technical assistants. Together with the UCT eResearch Centre and the ICT Department, they provide access to the datasets and tools that assist researchers in enhancing and completing their research (University of Cape Town Libraries, 2017).

During a presentation at NeDICC on the Digital Library Services at UCT, it was mentioned that the UCT RDM policy would be finalised in March/April 2016 (Mias, 2016). At the time of this study, however, the draft of the policy had not been approved by Council yet (University of Cape Town Libraries, 2017). In 2016, UCT also created their own version of DMPOnline, an online tool with funder specific templates for DMPs, which was originally created by the Digital Curation Centre in the UK (Mias, 2016).

At the end of 2016 and in the beginning of 2017, the RDM entity of UCT Libraries conducted an investigation on an institutional data repository for UCT (Roos and Mias, 2017). Potential software solutions were evaluated by using the following criteria: storage; operation and maintenance; publication workflow; dissemination and sharing; reporting, archiving and preservation; ingest; and visualization and analysis (Roos and Mias, 2017). A number of software solutions were identified by investigating which software solutions were used by leading research universities, and also by consulting the Re3Data website (Roos and Mias, 2017). The following software solutions were considered: DSpace, Fedora, Dataverse, Figshare, Dryad, Tind, CKAN, Zenodo, Globus and EPrints. It was found that the most popular software solutions were DSpace, Fedora, Dataverse and Figshare (Roos and Mias, 2017). The investigation further revealed that universities in the USA favoured Dataverse, those in Europe favoured DSpace and those in Australia favoured Figshare. Possible options were considered, such as commercial Software-as-a-Service (SaaS) (e.g. Figshare or Tind with cloud storage), local open source (e.g. Dspace or Dataverse or Fedora), a free online platform (e.g. Zenodo) and Hybrid (local SaaS) (e.g. Figshare or TIND installed on top of local storage).

During the eResearch Africa 2017 Conference held at UCT, representatives from various universities across South Africa held an open discussion with DIRISA and representatives from Figshare to discuss a possible roadmap for a South African Figshare consortium (eResearch Africa, 2017). This was followed by regional meetings that were arranged by DIRISA in Pretoria and Durban in July 2017. During these meetings, Figshare was introduced and demonstrated to research institutions and academics (DIRISA, n.d.). The outcome of this was a six-month trial period where institutions across South Africa could pilot Figshare at their institutions. In September 2017, the UCT Library completed the trial period and started the implementation of Figshare as a data repository at UCT (Zimmer, 2017). UCT also indicated that they had successfully implemented UCT communities within Zenodo. UCT is furthermore in the process of taking up Open Science Framework (OSF) services (Roos, Mias and Van Rooyen, 2017).

4.4.4.6 University Of Kwa-Zulu-Natal (UKZN)

Representatives from UKZN have been involved in the Africa Centre for Population Health. For more information on this Centre see 4.4.5.3.

4.4.4.7 University Of Pretoria (UP)

The Research Data initiatives at UP are discussed in more detail in 4.8.

4.4.4.8 University Of South Africa (UNISA)

The UNISA Library obtained a directive/mandate in 2011 from their University Executive, to investigate RDM at their university. The UNISA Library assembled a Library RDM Project Team, consisting of six members, who conducted a survey (situation analysis) among their researchers from March to July 2011, to determine the needs of researchers and the readiness of their institution for RDM (Lötter, 2014b; Macanda, Rammutloa and Bezuidenhout, 2014; Darries, 2016). After this, a RDM Task team was formed, which compiled high-level business requirements on RDM for the institution (Lotter, 2014b; Darries, 2016). The task team subsequently proposed a data management process flow (Macanda, Rammutloa and Bezuidenhout, 2014). At the time of this thesis, UNISA was investigating and testing DSpace as a possibility for a data repository, through a pilot study (Bezuidenhout, 2016). They are also in the process of developing an RDM strategy for the university in collaboration with the Research Department and other stakeholders (Bezuidenhout, 2016).

4.4.4.9 University Of The Western Cape

RDM Initiatives at UWC in 2016 consisted of the following:

- An investigation into the data practices of two research units at the institution, namely Bioinformatics and Poverty and Agrarian studies, which looked at aspects such as focus, storage, organisation, documentation, formats, loss of data, sharing practices, budgeting and the need for services;
- Training of library staff in RDM; and
- Engagement with others on campus, for example through meetings, a campus workshop and consultations (Fullard, 2016).

4.4.4.10 University Of The Witwatersrand (WITS)

At Wits University, the Library is taking the initiative with the provision of materials, documents and consultation/training on RDM to researchers, via a website. The Library also has a RDM consultant available to assist researchers (University of the Witwatersrand Library, n.d.). The university has furthermore conducted a research survey to determine the needs and readiness of their institution for RDM.

In 2016, the School of Computer Science and Applied Mathematics instituted a BSc Honours in Big Data Analytics, in response to needs of industry and the scientific community, where huge demand for skilled data scientists has arisen (University of the Witwatersrand, 2017a). The Wits School of Public Health also offers an MSC in Epidemiology in the field of Public Health Informatics, “which aims to develop research data management as a specialist qualification with reference to large longitudinal studies” (WITS, 2017b).

4.4.5 Other Initiatives

4.4.5.1 Southern African Large Telescope (SALT)

The Southern African Large Telescope (SALT) is “the largest single optical telescope in the southern hemisphere” and is located at the SAAO site in Sutherland, Western Cape, South Africa (Southern African Large Telescope, 2017). SALT has been in full science operation since 2011, and has been delivering a relatively small amount of data at around 5-10 GB per night (Crawford, 2013; Southern African Large Telescope, 2017). With the introduction of high-speed instruments like BVIT and new instruments like the NIR, it was expected that SALT would be able to produce up to 250GB per night of raw data (Crawford, 2013). In addition to SALT, the observing station in Sutherland also has a number of small remote and/or robotic telescopes (Crawford, 2013). “These include the three 1m telescopes for LCOGT, the Solaris telescopes searching for additional solar planets,” as well as KMTNet, which would have a wide field imager (Crawford, 2013). Individually, each of these telescopes should dispatch between 60-200 GB of data per night. In addition, upgrades to the small telescopes were also planned. Together, all the telescopes and facilities at Sutherland would be generating almost 1 TB of data per night (Crawford, 2013). Data products are archived by SAAO and are visible through the VO interface (https://vodasdata.salt.ac.za), but is only accessible to SALT partner astronomers (Cesarsky et al., 2015: 14).

4.4.5.2 Square Kilometre Array (SKA)

The SKA project is an international project to build the largest radio telescope in the world, and will be co-located in South Africa and Australia (SKA South Africa, n.d.(a)) The SKA project is overseen by a non-profit company, SKA Organisation, which has its headquarters at Jodrell Bank Observatory near Manchester, UK (SKA South Africa, n.d.(a)). The SKA Organisation was established in December 2011 "to formalise relationships between international partners and to centralise leadership of the project" (SKA South Africa, n.d.(a)). In South Africa, the Karoo Array Telescope (MeerKAT) is a radio telescope that was developed as a precursor to the SKA telescope, and the plan is to integrate this into the mid-frequency component of Phase 1 of the SKA (SKA South Africa, n.d.(b)). The SKA is expected to generate huge volumes of data and, as mentioned in 4.4.3.7, the SKA will involve large teams of international researchers with accompanying large data quantities, and powerful processing and analysis needs. To assist the SKA in managing the volumes of data that will be generated, IBM and the Netherlands Institute for Radio Astronomy (ASTRON) is collaborating on a project called DOME, with the aim of developing a computing and processing system for the SKA (Perry, 2013). Work on this project started in February 2012 and crucial research into technologies that will be needed to be built in the first half of the next decade, has been carried out since (ASTRON and IBM Center for Exascale Technology, 2017). Three main areas were focused on:

- Green computing, focusing on technologies that will "radically reduce the power" that will be required "to do computationally intensive work on extremely large" volumes of data;
- Data and streaming, focusing on technologies that will be needed "to process data on-the-fly" and store these data at a high efficiency for later use; and
- Nano-Photonics, focusing on technologies that will be needed to drastically decrease the power needed for data transport over long distances and within computing machines (ASTRON and IBM Center for Exascale Technology, 2017).

The NRF has joined the collaboration as a user platform member. The DOME collaboration will enable a partnership between scientists and engineers from public and private institutions, and could lay the groundwork for the scientific community to solve other data challenges such as climate change, genetic information and personal medical data, according to Simon Ratcliff (technical coordinator of DOME South Africa) (Perry, 2013). Researchers from UP have also been involved in this project (Smit, 2015).

The SKA project could potentially help with the management of big data sets. In 2016, the Netherlands and South Africa set up a data science partnership to develop and establish national and regional data centres to address the challenges of managing, processing and making accessible the vast volume of data that the SKA project will generate (Netherlands Organisation for Scientific Research, 2016). These data centres will enable researchers across the world to access large-scale data infrastructures and high-performance computing that will be required to make sense of the data (Netherlands Organisation for Scientific Research, 2016). The partnership consists of the Netherlands Organisation for Scientific Research (NWO), ASTRON, IBM, SKA South Africa and UNISA, and their “ground-breaking research project” is called ‘Precursor Regional Science Data Centres for the SKA’ (SKA-RSDC) (Netherlands Organisation for Scientific Research, 2016). The topic of big data is elaborated upon in 4.6 of this study.

4.4.5.3 The Africa Centre For Population Health

In 1996, the Wellcome Trust in the UK sent out a call for applications to establish an international centre for population and reproductive health in sub-Saharan Africa (Africa Centre for Population Health, 2015). The University of Natal and Durban Westville (later University of KwaZulu-Natal) and the South African Medical Research Council collaborated to form the Centre, which was at that stage named the Africa Centre for Population and Reproductive Health (Africa Centre for Population Health, 2015). The Centre was renamed in 2002, to become known as the Africa Centre for Health and Population Studies, in order to more specifically reflect the expanse of health and population research being done (Africa Centre for Population Health, 2015). The Centre underwent a transformation in 2015 and became known as the Africa Centre for Population Health. The centre is managed by a Governance Committee, which includes representatives from UKZN, University College London, and the Wellcome Trust (Africa Centre for Population Health, 2015).

The Centre has a data management platform that links population research, clinical research and lab research. The platform presents a favourable research environment for researchers, increases access to high quality research data by scientists, renders help to researchers in formulating and implementing DMPs, and makes it possible to assess and certify the quality of all research data gathered at the Centre (Africa Centre for Population Health, 2015). Data is captured from three research platforms by using a number of tools (Africa Centre for Population Health, 2015). Data from the Population Research Platform is lifted using the ACDIS dot Net application. It is then stored in the Demographic Information System Database (ACDIS) (Africa Centre for Population Health, 2015). HIV data originating from the Clinical Research Platform is laid hold off by using the Tier dot Net application, and is stored in the clinical database (ACCD) (Africa Centre for Population Health, 2015). The Laboratory Information Management System (LIMS) is applied in the laboratory for the gathering and logging of samples and related data. The system is also set to use electronic data gathering and capturing application tools such as Redcap and Open Data Kit (ODK) (Africa Centre for Population Health, 2015). ACDIS Util and ACDIS VA are used to do quality assurance on the integrated database (Africa Centre for Population Health, 2015).

4.4.6 Potential Partners In RDM In South Africa

4.4.6.1 Southern African Research And Information Management Association (SARIMA)

SARIMA is an outgrowth of the Research Directors Forum (RDF) and was formally established as a stakeholder organisation in Cape Town on 14 February 2002 (SARIMA, 2014). It “operates at an institutional, national and international level, as well as across the research and innovation value chain, from research management to successful innovation (commercialization)” (SARIMA, n.d.). The main function of SARIMA is “to promote research and innovation management for the benefit of southern Africa” (SARIMA, n.d.). Its objectives are:

- Professional development and capacity building in managing research and/or innovation;
- Promotion of best practice in the management, administration and support of research and innovation;

- The creation of awareness in academic and public forums of the value of "a more robust research and innovation system and the benefaction of it to socio- and economic development";
- Promotion of pertinent national and institutional policy that encourages research and innovation;
- Participation in the development and testing of policy;
- Advancement of science, technology and innovation, which includes dealing with imbalances in access to, and flowing of, knowledge between the North and the South; and
- Spearheading research and innovation management enhancement within southern Africa, which includes guidelines for the varied components of the research and innovation cycle (SARIMA, 2014).

SARIMA, in collaboration with the DST, are responsible for the management of a number of Southern African Development Community (SADC) focal points to advance research and innovation management. In addition, SARIMA manages and co-ordinates a growing list of multilateral programmes and projects in support of the objectives of SARIMA and to the advantage of its members and stakeholders (SARIMA, n.d.). SARIMA could also potentially play a role in providing guidelines or informing members on issues of RDM as part of the research and innovation cycle.

RDM in South Africa originally developed in a haphazard manner, with some institutions such as the HSRC already starting RDM as early as 2007 and UP releasing a policy on retention of data in 2007, as well as the establishment of SADA by the NRF in 2009. It was only with the establishment in 2010 of NeDICC, representing the data curation community, the establishment of the government initiative DIRISA (one of the initiatives of NICIS), the recommendation of South Africa as one of the sites of the SKA in 2012, as well as the release of the NRF Statement on Open Access in 2015, that efforts for a collective coordinated approach for RDM in South Africa gained momentum. Some of the higher education institutions have been putting much effort into the development of RDM policies for their institutions, and many of these institutions have also created websites or 'libguides' to assist their researchers with RDM-related issues. Currently, many of the research institutions are using the DMP Tool that is available on DIRISA's website, and a number of these institutions are also participating in piloting Figshare, to test if it will be a suitable solution for a data repository. There is also a great need for skills development and capacity building to address the needs around RDM, with some of the research institutions starting to implement courses to address this. NeDICC has also been used as a forum for capacity building, but unfortunately does not have the authority to influence national policy.

Much more work still needs to be done to establish RDM practices, infrastructure and policies in all the research institutions across the country. In addition, RDM developments in South Africa have tended to focus on only certain components of the research data lifecycle and not on the whole research data lifecycle. The next section will give an overview of the research data lifecycle.

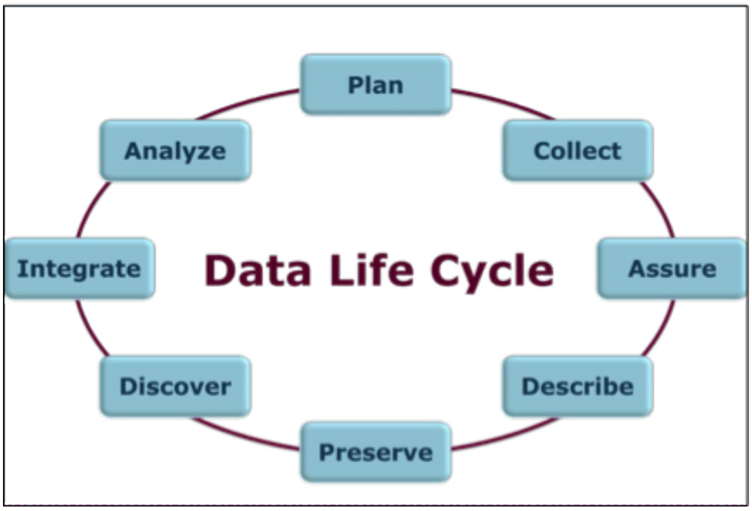
4.5 RESEARCH DATA LIFECYCLE

4.5.1 Introduction

A review of literature (Ball, 2012; Beagrie, 2004; DCC, 2018a; Wiggins et al., 2013: 1-14) disclosed that research data typically flow through a data lifecycle. An array of different data lifecycles are discussed in literature, for example DataOne's lifecycle (Wiggins et al., 2013: 1-14), Data Documentation Initiative's Combined Lifecycle Model (Ball, 2012: 7), and the DCC Curation Lifecycle Model (DCC, 2018a), but for the purposes of their article, the authors chose to use the UK Data Archive Lifecycle (UK Data Archive, 2014), as many of the components in the other data lifecycles are subsumed in this lifecycle; it relates well with the research lifecycle; and also clearly shows distinct data management actions that could be taken in each stage.

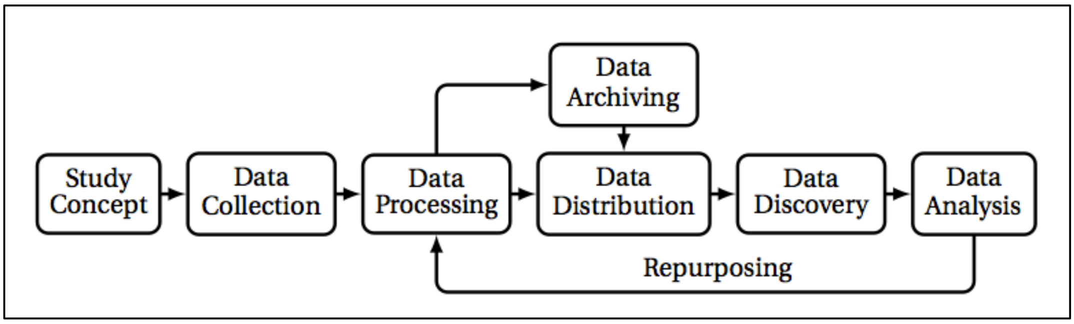
DataOne suggests a data lifecycle (see Figure 4.4) with the following components (steps): plan, collect, assure, describe, preserve, discover, integrate, analyse (Wiggins et al., 2013: 1-14). These steps can take place in any number of different sequences, with some occurring simultaneously and some repeated more than once (Wiggins et al., 2013: 2). The DataOne cycle focuses more on the processes within a lifecycle than the stages themselves, and was therefore not deemed a suitable cycle for this study. Its components, however, were deemed valuable.

Figure 4.4: DataOne Data Lifecycle



The Data Documentation Initiative (DDI) version 3.0 Conceptual Model [Str04] includes a Combined Lifecycle Model (See Figure 4.5) for research data, particularly social science data. This model is mostly linear, with one alternative path and one feedback loop. The model contains the following sequential elements: study concept, data collection, data processing, data distribution, data discovery, data analysis, with a feedback loop from data analysis to data processing called repurposing, and an alternative path from data processing to data archiving, and then sequentially to data distribution (Ball, 2012: 7). This model, though very useful, was deemed too linear, and not really a research cycle, but the elements within the model were considered valuable for this study.

Figure 4.5: Data Documentation Initiative's Combined Lifecycle Model



The Digital Curation Centre suggests a more comprehensive model (See Figure 4.6), which they call the DCC Lifecycle Model, (DCC, 2014b). This model has data (which includes digital objects and databases) in its centre. Around this are the full lifecycle actions, which include description and representation information, preservation planning, community watch and participation, and curate and preserve. Sequential actions of the cycle include: conceptualise, create or receive, appraise and select, ingest, preservation action, store, access use and re-use, and transform. Occasional actions include dispose, reappraise, and migrate. Although this model is very comprehensive, the author of this thesis found that the model does not cover the full spectrum of all the stages of a research data lifecycle. For example, the stages of processing data and analysis of data are not clearly demarcated/covered.

The UK Data Archive proposes a research data lifecycle (See Figure 4.7) that consists of sequential stages that starts at the creation of data by researchers, the processing of the data, the analysis of the data, the preservation of the data, giving others access to the data, and the re-use of data by other researchers (UK Data Archive, 2014).

The relations and contrasts between components of the mentioned data lifecycles are summarised in Table 4.3.

Figure 4.6: DCC Curation Lifecycle Model

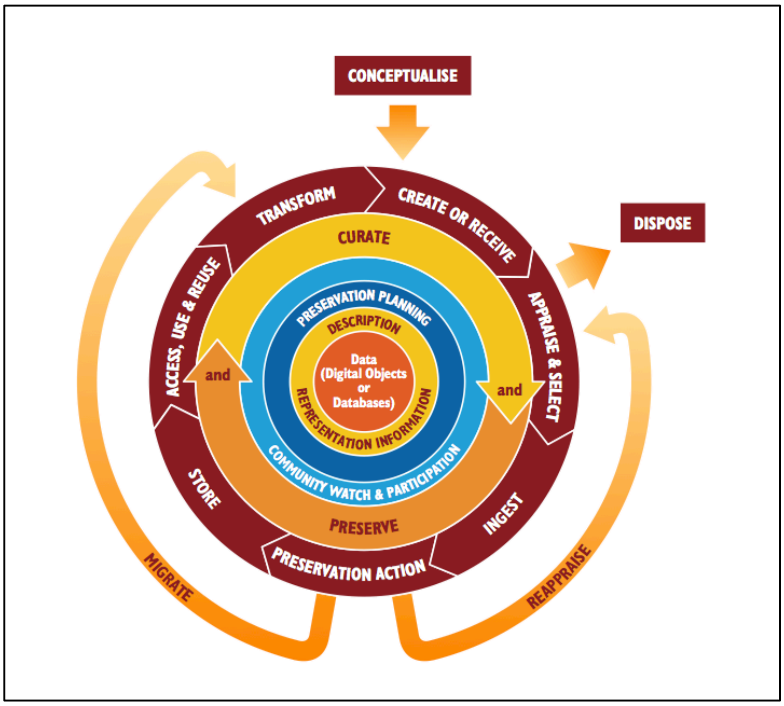
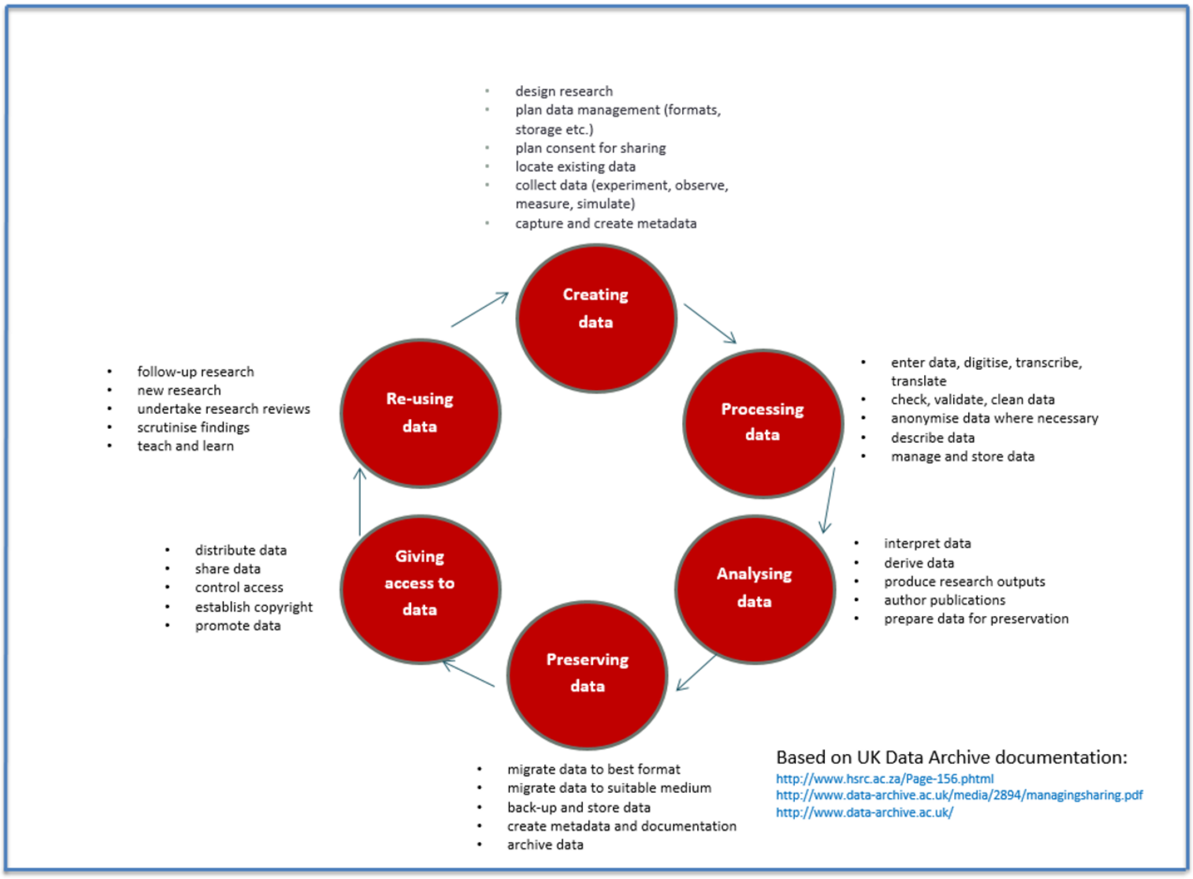


Table 4.3 shows that the stages of the UK Data Archive can encompass and draw together the major components of the other data lifecycles. The action taking place in the whole lifecycle as mentioned in the DCC cycle was also integrated with the cycle from the UK Data Archive. This cycle was then adapted for the purposes of this study (see Figure 4.7).

Table 4.3: Comparison Of The Components Of The UK Data Archive Lifecycle, Dataone Lifecycle, DDI Version 3.0 Conceptual Model, And DCC Lifecycle

UK Data Archive	DataOne	Data Documentation Initiative (DDI)	Digital Curation Centre (DCC)
Creation of data	Plan	Study concept	Conceptualise
	Collect	Data collection	Create or receive
Processing of data	Assure	Data processing	
Analysis of data	Analyse	Data processing	Appraise and select
	Discover		Description and representation information
Preservation of data	Describe	Data archiving	Preservation action
	Preserve		Ingest
			Store
			Curate and preserve
Giving others access to data	Discover	Data distribution	Access use and re-use
		Data discovery	
		Data analysis	
Re-use of data	Integrate	Repurposing	Transform
			Dispose
			Re-appraise
			Migrate
Whole Cycle			Community watch and participation
			Preservation planning

Figure 4.7: Research Data Lifecycle (Adapted From UK Data Archive Cycle)



Based on UK Data Archive documentation:
http://www.hsrc.ac.za/Page-156.phtml
http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
http://www.data-archive.ac.uk/

4.5.2 Stages Of The Research Data Lifecycle

In each of the stages of the research data cycle, various RDM actions can be taken to ensure the value, quality, accuracy, accessibility, long-term availability, intelligibility, and security of data. These actions form the essential components of a research data lifecycle.

4.5.2.1 Creating Data Stage

(a) Designing Data Management Plans

In this stage, researchers design DMPs. A DMP is described as “a formal document that outlines what you will do with your data during and after you complete your research” (University of Virginia Library, 2014). Tools that can assist researchers in their data management planning are: the Data Management Planning Tool (DMPTool), available at https://dmptool.org/ from the University of California Curation Center of the California Digital Library, as well as the DMPonline tool, available at https://dmponline.dcc.ac.uk/tool, from the Digital Curation Centre in the United Kingdom. Each of the funders provide requirements with regard to DMPs, while the Research Office of each university could provide researchers with information on funders' DMP requirements and assist in applications to funders. Librarians can also play an important role to train and advise researchers on the different DMP tools that are available.

(b) Data Capture / Collection

This is the action or process researchers employ to gather and measure "information on variables of interest, in an established systematic fashion" that will enable them "to answer stated research questions, test hypotheses, and evaluate outcomes" (Northern Illinois University, n.d.; The Oxford Dictionary, 2014). Various data collection methods can be employed, for example observations, textual or visual analysis, interviews, focus group interviews, surveys, tracking, experiments, case studies, literature reviews, questionnaires, etc. (Gill et al., 2008: 291; Mack et al., 2005: 2-3; Onwuegbuzie, Leech and Collins, 2012: 1-28; Quantitative Data: Surveys, 2014; Yin, 2009, 1-240). Other researchers' data can also be re-used for further research (Corti et al., 2014: 2). In addition, data could be captured from instruments with tools such as MyTardis, and moved into an environment where computations can be done with it, or into an archive for storage or into a repository to be published (About MyTardis, n.d.). A VRE Manager and / or the VRE Champion can play an important role in advising researchers in the various data collections methods and data capture tools and see to it that these are captured in an organised manner onto a VRE. Support can be rendered by the University IT department with regards to various data capture equipment, instruments and infrastructure, by integrating these into a VRE framework, and providing consultation and training.

(c) Data Storage

The University of Wisconsin Madison describes storage as the preservation of data files in a secure location, which can be accessed readily. They differentiate between storage and backups, and describe the process of backups as the preservation of additional copies of your data in a separate physical location from data files in storage (Research Data Services, University of Wisconsin-Madison, 2014). Data storage and backups can be done in the data creation stage, the processing data stage, as well as in the data analysis stage. Researchers in a VRE typically take primary responsibility for storing their data files on a VRE system. The University IT department takes responsibility for the installation, availability and maintenance of data storage infrastructure in a VRE, while the University Executive should take responsibility for the provision of the necessary resources (e.g. funding and staff), for data storage infrastructure management. The Librarian can provide training and consultation to researchers on file naming conventions for data storage in a VRE.

(d) Metadata Creation

The USGS defines metadata as “information that describes a dataset, such that a dataset can be understood, re-used, and integrated with other datasets” (USGS Data Management, n.d.). Metadata, according to Corti et al. (2014: 38) provide information that is searchable, standardised, and structured. This information explains “the aim, origin, time references, geographic location, creating author, access conditions and terms of use of a data set.” Metadata also helps researchers in locating existing data resources, while providing a bibliographic for citing the data (Corti et al., 2014: 38). Researchers typically should take responsibility for adding metadata to the data files. Librarians have the skillset to provide consultation and training on metadata schemas that are stored on a VRE and uploaded onto a data repository. The VRE Manager and / or the VRE Champion should also monitor the adding of metadata by researchers.

4.5.2.2 Processing Data Stage

Data processing can consist of:

(a) Data Cleansing

Data cleansing is the process of detecting errors and inconsistencies in data, and then correcting, replacing or removing these in order to enhance the quality of the data. (Rahm and Do, n.d.; Sarpong and Arthur, 2013: 14). In a VRE, this process is the primary responsibility of the researcher, but the VRE Manager and / or VRE Champion should also monitor and facilitate this. Librarians could consult and train researchers in the various tools that are available to do data cleansing, e.g. OpenRefine and Trifacta.

(b) Data Validation

Data validation is the process "to determine if data quality goals have been achieved and the reasons for any deviations. Validation checks that the data makes sense" (Martin and Ballard, 2010: 8; United States Environmental Protection Agency, 2002: 15). Data Validation in a VRE is primarily the researcher's responsibility, with monitoring from the VRE Manager and / or VRE Champion.

(c) Data Anonymisation

Data anonymisation is "the process of de-identifying sensitive data, while preserving its format and data type" (Raghunathan, 2013: 4). Cormode and Srivastava (2009), Raghunathan (2013: 172-182), and Vinogradov and Pastsyak (2012: 163) suggest a number of anonymisation techniques:

- Suppression, which concerns the removal of information (e.g. gender) from the data;
- Generalisation, where information (e.g. age) is made common or unrefined, for example changing them into sets such as age ranges;
- Perturbation, which is a statistical-based method that "entails the protection of confidential/sensitive data by adding random 'noise' to confidential attributes in the data, thereby protecting the original data" (Wilson and Rosen, 2003: 15);
- Permutation, where sensitive associations between entities (e.g. purchase of medication by a person) are swapped (Cormode and Srivastava, 2009);
- Substitution, where identifiable numbers or contents of a data column are replaced with data from a predefined list of fictitious but similar data types so it cannot be traced to the original subject (Charles, 2012);
- Shuffling, where the data is shuffled in one column, for example the combination (name, bank account) will not be real;

- Number and date variance, which “involves modifying each value in a column by some percentage of its real value to significantly alter the data to an untraceable level” (Charles, 2012);
- Nulling out, where sensitive data is simply removed, by deleting it from the shared data set, and replaced with null values (Charles, 2012);
- Data masking**,** where sensitive data is rendered unintelligible by replacing it with other data — usually characters that will meet the requirements of a system designed to test or still work with the masked results. Masking safeguards vital parts of personal identifying information, for example the first five digits of an identification number are obscured or otherwise de-identified (Simpson, n.d.); and
- Data encryption techniques, where data is converted and transformed “into scrambled, often unreadable, cipher-text using non-readable mathematical calculations and algorithms.” The data can then only be read by using a “corresponding algorithm and the original encryption key” (Simpson, n.d.).

Data anonymization in a VRE is primarily the researcher’s responsibility, with monitoring from the VRE Manager and / or VRE Champion.

4.5.2.3 Analysing Data Stage

(a) Data Interpretation And Analysis

Data interpretation and analysis “is the process of assigning meaning” to the gathered information and ascertaining “the conclusions, significance, and implications of the findings” (Analyzing and Interpreting Data, n.d.). In a VRE, the researcher takes responsibility for data interpretation and analysis, with support and guidance from the VRE Manager and / or VRE Champion. Librarians could provide the necessary training and consultation on data analysis tools, e.g. R, Stata, SPSS, etc.

(b) Data Publishing

Data publishing is the process of making research data underpinning the findings published in peer-reviewed articles, available for readers and reviewers in an appropriate repository, or “as supplementary materials to a journal publication” (Corti et al. 2014: 197). A more recent development has been the appearance of data journals. These journals publish data papers that describe a dataset, and also give an indication in which repository the dataset is available (Corti et al. 2014: 7-8). In a VRE, the researcher takes primary responsibility for the publishing of his / her data. Librarians could provide guidance on appropriate journal publications, and could provide the necessary training and consultation on publishing in a data repository. The VRE Manager and / or VRE Champion could monitor the process. Peer reviewers could also be given access to a VRE and some of its components, e.g. a repository, to ensure that the published data is of a good quality.

(c) Data Visualisation

Data visualisation is described by Friendly (2009: 2) and Schnell and Shetterley (2013: 3) as the visual representation of data, and is used to enable people to both understand and communicate information through graphical and schematic avenues. Schnell and Shetterley (2013: 3) further differentiate between exploratory and explanatory data visualisation. Exploratory visualisation is used to explore and make sense of data, while explanatory visualisation is used to explain and communicate a finding after the analysis of the data is complete (Schnell and Shetterley, 2013: 3). In a VRE, researchers can use visualisations as part of their data analysis. Librarians could provide researchers with consultation and training in visualisation software, e.g. Zoho Reports, Google Fusion Tables, etc.

4.5.2.4 Preserving Data Stage

(a) Data Archiving

As discussed earlier in this chapter, data archiving can be described as the process of retention and storage of valuable data for long-term preservation, so that the data will be protected from risk (i.e. loss, or corruption) and will be accessible for future use. In a VRE, it is primarily the responsibility of the researcher to ensure that data are archived. The VRE Manager and / or VRE Champion should monitor that this is done.

(b) Data Preservation

The Data Conservancy community, headquartered at the Sheridan Libraries at Johns Hopkins University, describes data preservation as the process of providing enough representation information, context, metadata, fixity, etc. to the data so that anyone other than the original data creator can use and interpret the data (Ruth Duerr as quoted by Choudhury, 2014: 125). In a VRE, this is an area where the librarian could assist, as well as the University IT, to provide the necessary preservation metadata, and creation of checksums etc.

(c) Long-Term Data Preservation

Long-term data preservation is defined by the University Library and the University Computing Service at the University of Cambridge as "the process of maintaining data over time so that they can still be found, understood, accessed, and used in the future" (Cambridge University Library, 2012). In a VRE, the University IT would need to maintain the IT infrastructure, e.g. storage systems needed for long-term preservation of data. The librarian could provide the necessary consultation on the formats for long-term preservation, e.g. tiff, pdf, etc.

(d) Linking Data To Research Outputs

This is the process of connecting the underlying data relating to a specific research output, e.g. journal article, thesis, etc. to the research output itself. This could be done by adding a digital object identifier (DOI) to the dataset and including this in the metadata of the research output, or by citing the dataset (Callaghan et al., 2013). In a VRE, the researcher could add a DOI to the dataset, or a librarian could assist. Librarians can also provide the necessary consultation and training on DOIs and data citations.

4.5.2.5 Giving Access To Data Stage: Data Sharing

Data sharing is the process of opening up access to research data and making it available to other researchers (Corti et al., 2014: 2). Data sharing can be viewed as a valuable component of the scientific process. The sharing of data affords "opportunities for other researchers to review, confirm or challenge research findings” (USA Department of Education, Institute of Education Sciences, n.d.). Data sharing can also improve and augment scientific inquiry through a range of other analytic endeavours, including the use of shared data to: test alternative theories or hypotheses; explore alternative sets of research questions than those targeted by the original researchers; combine data from multiple sources to obtain potential new insights and areas of inquiry; and/or conduct methodological studies to improve research methods and statistical analyses (USA Department of Education, Institute of Education Sciences, n.d.). In some cases, data can be restricted because of confidentiality, legal or commercial reasons.

Different levels of data set confidentiality are differentiated in the Privacy Regulations released under the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) (United States of America, 1996). Morse et al. (2011: 1) list these as:

- **A protected data set**: This is most confidential data set that includes protected information that can unequivocally identify a person (in case of health data this can only be shared with a patient’s care-givers, e.g. doctors, and the institution that is providing the care);
- **A limited data set**: This type of data set does not include information such as names, addresses, identity numbers or medical insurance numbers, but could include things such as geographical region, birthdate, and in the case of health information, the date admitted in hospital, as well as dates interviewed, etc. A limited data set can be shared with a research group, with a written legal agreement;
- **A de-identified data set**: This is the least confidential type of data set. It has all the data that can identify a person, removed from it, for example, names, dates, and more specific, geographical information such as the address and telephone numbers. Two types of processes can be followed to de-identify a data set. These are “the reversible process” of de-identification, where “the data are key-coded, encrypted, or pseudonymized to remove personal information”, and the “irreversible process” of anonymisation, where “data are completely stripped of all identifying information that can be linked to the study participants” (National Research Council, 2010).

Developments in technology, however, have made it easier for intruders/hackers to determine a person’s identity from data that have been de-identified or anonymised. These confidentiality risks, according to National Research Council (2010), consist of two types: identification disclosure risk and attribute disclosure risk. Identification disclosure takes place when an intruder establishes that information on a selected individual can be found in a specific data file. It might then be possible for the intruder/hacker to ascertain which of the records in the file belongs to the specific researcher, by scrutinising the demographics or other variables (National Research Council, 2010). Attribute disclosure happens when an intruder finds out what the value of a sensitive variable is in relation to a specific person, which might enable him/her to identify records belonging to that person.

In some instances, the data can be subject to legal restrictions, where the data and / or dataset carries a copyright license or a Creative Commons license that determines under which circumstances and specifications the data may be shared or used (Creative Commons, n.d.). Data and datasets may also be restricted because of commercial reasons. Data could, for example, be restricted during the registration of a patent, or could contain information that includes intellectual property rights.

In a VRE, librarians could provide the necessary consultation and training on copyright licenses, e.g. Creative Commons, and intellectual property rights. The VRE Manager could also provide advice and direction with regards to ethical processes and intellectual property rights. In this stage, members of the community could be given read-only access to certain parts of the data, for example data in repositories, wikis or blogs.

4.5.2.6 Re-Using Data Stage

(a) Data Re-purposing / Re-use

This is the process where secondary data (data that have been captured and analysed by other researchers) can be re-analysed, re-worked or used for new analyses, and compared with contemporary data. This process “also enables research where the required data may be expensive, difficult or impossible to collect,” e.g. large-scale surveys, or historic data (Corti et al., 2014: 169). In a VRE, librarians could assist researchers to find datasets that are related to the research topic they want to investigate (They could provide the necessary consultation and training to researchers, or conduct these searches on behalf of researchers). The researcher can search for these datasets themselves or consult a librarian, and then use these for new analyses.

(b) Data Citation

Data citation is the process of referencing (attributing and acknowledging) re-used data in a similar fashion as traditional sources of information (Corti et al. 2014: 197). Citing a data collection acknowledges the author's sources, helps in identifying and finding the data, encourages the reproduction of research results, enables the tracking of the impact of data, and provides a framework for the recognition and rewarding of data authors (Corti et al., 2014: 205). Processes for citing data have taken some time to develop. In 2009, the American Psychological Association included guidance on citing data in the Publication Manual of the American Psychological Association, followed by Oxford University Press in their Oxford Style Manual (American Psychological Association, 2009; The New Oxford Style Manual, 2012). In 2007, Altman and King proposed a robust citation of data in the social sciences. They argued for six components: author(s), title and publishing date of the dataset, as well as a unique global identifier, a universal numeric fingerprint, and a bridge service, which would persist and identify the data even when the publishing technology or location changes (Altman and King, 2007). To enable a unique and persistent identification of a digital document on the Internet requires the use of a Uniform Resource Name (URN). Three types of URN issuing and resolving services have developed: the Digital Object Identifier (DOI) system, the URN: NBN system developed by the Germans, and the Archival Resource Key (ARK) system developed by the California Digital Library. The DOI is the one service that is best known, and is described by Corti et al. (2014: 206) as "a string of characters that make up a digital identifier, which is used to uniquely identify an object." Metadata about the object is stored with the DOI name, and can contain a web address (a uniform resource locator - URL), indicating where the object can be found (Corti et al., 2014: 206). In a VRE, the primary responsibility for data citation lies with the researcher. Librarians could also provide the necessary consultation and training on DOIs and data citations.

4.5.2.7 Processes / Actions Taking Place Through The Whole Research Data Lifecycle: Data Provenance

The provenance of data is an essential feature in management of data and should be considered in any RDM initiative. Data Provenance can be described as the "history of a data file or data set, including collection, transformations, quality control, analyses, or editing" (Strasser et al., 2012: 11). Such data would typically include information on the person(s) responsible for the data set throughout its lifetime, the context of the data set with regard to a larger project or study, as well as revision history, including additions of new data and error corrections (Strasser et al., 2012: 7).

In a VRE the VRE designer would typically be involved in all the stages of the research data lifecycle to develop, monitor, maintain and adapt the VRE to researchers' needs.

4.6 THE MANAGEMENT OF BIG DATA

In 4.4.5.2, it was mentioned that the SKA project will generate huge volumes of data that will need to be managed. The management of big data, however, is a much more complex process than the management of small data sets, and necessitates a whole different approach. Big data "denotes those datasets which cannot be acquired, managed or processed on common devices within an acceptable time" (Huadong, 2014), and is further characterised by volume, velocity, variety and veracity:

- **Volume:** Concerns the size of the data sets that systems must ingest, process and disseminate. With big data, these data sets are huge and cannot be stored or analysed by conventional hardware and software;
- **Velocity:** This refers to the speed with which data is created, in other words, the pace at which data flows in and out from sources like business processes, machines, networks, sensors and human interaction with things like social media sites or mobile devices;
- **Variety:** Refers to the complexity of the types of information handled ("many sources and types of data both structured and unstructured"); and
- **Veracity:** "Refers to the biases, noise and abnormality in data", which increase in big data (Normandieu, 2013).

In the academic sphere, Big Data is creating an all-new approach to research, which Jim Gray announced as the ‘Emergence of a Fourth Research Paradigm’ during a talk in 2007, which he called Data-Intensive Science (where researchers are challenged with data sets from many different sources, e.g. captured by instruments, generated by simulations or generated by sensor networks) (Jim Gray on eScience: a transformed scientific method, 2009**)**. Data-intensive science, according to Bell (2009: xv), consists of three basic activities: data capture, data curation and data analysis, and suggests that a generic collection of tools is needed. These tools should contain the full range of activities from capture and data validation through curation, analysis, and ultimately permanent archiving/preservation. Curation is about “finding the right data structures to map into various data stores, which includes the schema and necessary metadata for longevity and for integration across instruments, experiments and laboratories” (Bell, 2009: xv). Data analysis incorporates a whole series of activities throughout the workflow pipeline, inter alia the use of databases (as an alternative to a set of flat files that a database can access), analysis and modelling, and finally, visualisation (Bell, 2009: xv).

Big data is not something that can be managed by one institution alone, with the result that there are various international big data initiatives running in a wide range of disciplines. Researchers and higher education institutions should avail themselves of the various initiatives. Memberships to task groups and workgroups of organisations such as CODATA, the RDA, International Federation of Data Organisations for Social Science, and the World Data System, will also add to the expertise and know-how in managing Big Data.

4.7 THE VALUE OF RDM

RDM can be of considerable value to a higher education institution, as well as to researchers. The majority of universities (as can be seen through an examination of a number of university library websites, e.g. Columbia University Libraries, n.d.; Concordia University, 2017; Durham University, n.d.; UCD Library, 2014; University of Edinburgh, 2014; University of Manchester Library, n.d.; University of Virginia Library, 2014) provide reasons for conducting RDM. These reasons are mostly similar. By combining the reasons given on these universities’ websites, a list can be compiled of reasons for managing research data. Managing research data will effectively:

- Enhance research practice and efficiency;
- Assist in addressing funding body grant requirements/mandates;
- Enable institutions and researchers to meet journal publisher requirements;
- Ensure accountability;
- Ensure research integrity, validation and replication;
- Ensure research data and records are accurate, complete, authentic and reliable;
- Increase the visibility and impact (citation rates) of research;
- Enable the preservation (storage and archiving) of data for medium and long-term;
- Save time and resources in the long run;
- Enhance data security and minimise the risk of data loss;
- Promote sharing of data and prevent duplication of effort by enabling others to use the data;
- Easy discovery of research data through the use of metadata, tagging, linking and search functionalities;
- Enable others to use your data, resulting in the reinforcement of scientific enquiry and new and unanticipated discoveries;
- Assist in complying with practices conducted in industry and commerce;
- Prevent unauthorised use by addressing privacy and confidentiality issues through and beyond the research project; and
- Foster international research collaborations.

Additional reasons provided in the literature are:

- Minimising the risk of legal challenges (e.g. with regards to validation of results, and intellectual property); and
- Protecting the institution from reputational, financial and legal risks (Ashley, 2012: 156; Giesen, 2015: 9)

4.8 RDM AT THE UNIVERSITY OF PRETORIA (UP)

4.8.1 First Initiatives

The first initiatives on RDM at UP came in 2007 in the form of a policy document, titled "Policy for the preservation and retention of research data" (Rt306/07) (Crew, 2007). This was followed by a survey of RDM practices at the University over the period October 2009 – March 2010, conducted by the Department of Library Services, with the aim to capture and document RDM practices across the University, and to identify problems and possible solutions (Pienaar, 2010). This project concluded that RDM does not exist in any formal manner (with the exception of one or two departments) at UP, and recommended that the DST's DIRISA should support UP's RDM needs. Secondly, a formal staff position of 'research data manager' should be created at UP (Pienaar, 2010).

4.8.2 Survey – August-October 2013

After deliberation with the Vice Principal Research and Postgraduate Studies, a second survey was conducted from August-October 2013 by the UP Department of Library Services, among the Deputy Deans: Research of all the faculties, to determine what is seen as essential data that the University should manage (Van Wyk, 2013a). This survey showed the following:

- **Essential data** at UP was identified as interview data, questionnaires, spreadsheet data, lab notebooks, experiment/laboratory data, images (e.g. graphs, models, sketches, X-Rays, scans etc.), literature reviews, sequencing data, and computer-generated data;
- **Level of data** that should be managed, included only raw data in some faculties, only processed/analysed data in others, and both raw and processed/analysed data in some faculties;
- **Volume of data -** A small number of faculties worked only with small data sets. The majority of faculties worked with small and big data sets, with an exponential increase in big data sets, which presented a challenge with regards to the provision of the necessary IT infrastructure that would be able to handle this;

- **The data formats** that were being used varied, and included Excel, Pdf, MS Word, Text Format, images in various formats, video, sound, various computer-generated formats, SPSS, SAS, AMOS, Qualtrics data, SurveyMonkey data, simulation data formats, and even data from social media. Some of the data were still in paper format, and a decision had to be taken about digitising these, or the provision of good storage facilities for long-term storage;
- **RDM Plans** - Most of the faculties had an internal arrangement of what should be done with regards to students' and researchers' data, but none of the faculties had a RDM plan in place. It was found that guidance and training would be needed on how to set up these plans;
- **Metadata** - The majority of the faculties indicated that they had no metadata in place, but there were some pockets of researchers that had metadata schemes in place. Big data sets however can be very complex and presented a huge challenge with regards to adding metadata to the millions of data points within these sets;
- **Uploading capacity** - None of the Faculties had any human capacity to upload the data sets to a repository; and
- **Willingness to share data** - The majority of faculties indicated that they would be willing to share their data under certain conditions, but one faculty indicated that they would not be willing to share their data, except in cases where they worked in a consortium where data were shared with other researchers in the consortium (Van Wyk, 2013a).

4.8.3 Pilot Projects

The Department of Library Services conducted five RDM pilot projects starting in 2013, and these were still ongoing at the time of the finalisation of this thesis. The first two projects were chosen by this researcher as case studies for this study.

4.8.4 Appointment of Assistant Director RDM

The next stage was the secondment of a senior staff member of the Department of Library Services (the researcher of this study) to the position of Assistant Director RDM in January 2014, with the directive to facilitate RDM at UP.

4.8.5 High Level Report on RDM

In July 2014, the Assistant Director RDM compiled a high-level report on RDM for the University Executive. This report included an overview of RDM, internationally and nationally, as well as the results from the survey on Essential Data that should be managed at UP (August-October 2013), with recommendations on the way forward (Van Wyk, 2014a)

4.8.6 Visit To Purdue University

During April 2014, the Deputy Director Innovation and Technology of the Library Services as well as the Library IT Specialist responsible for RDM, visited Purdue University in the USA to look at their Research Data Repository (PURR), as well as their long-term preservation processes, as a possibility for replication at UP (Van Wyk, 2014a). The idea was to establish a campus-wide database/repository for open access data sets with DOI's, similar to PURR. With regards to long-term preservation, these library colleagues were apprised of a hierarchical file packaging format called BagIt (Van Wyk, 2014a). BagIt is an outgrowth of work done by the Library of Congress in the USA, and has been widely adopted as packaging format by entities such as the Library of Congress, Archivematica digital preservation platform, Ghent University, Dryad scientific data repository, Stanford Digital Repository, and Central Connecticut State University (deposits bags on Amazon S3). In brief, the format entails that a virtual bag is created containing a data-set, as well as accompanying descriptive metadata (e.g. Dublin Core and MODS) and descriptive text files used to describe the contents of the bag. BagIt bags adds another set of metadata to an existing set of metadata, called the Preservation Metadata, e.g. PREMIS. This data is crucial for long-term preservation, which will aid the reconstruction of data at a later stage. It must be clear that stored bags and access to the bags must be located on secure and restricted infrastructure to avoid tampering and to safeguard the validity and integrity of the bags (Van Wyk 2014a). Although the creation and encoding of the bag is an automated process, there is a multitude of administration and processes that must be put in place to ensure proper preservation (Van Wyk, 2014a).

4.8.7 New RDM Policy For UP

During August 2014, the first draft of a new proposed RDM policy was compiled by the Assistant Director RDM (Van Wyk 2014b). This draft went through a number of iterations and the final draft was sent through to the University Executive for approval in January 2017 (Van Wyk, 2017). A number of shortcomings were identified in the original 'Policy for the preservation and retention of research data'; these were addressed in the new RDM policy:

- Clarification of concepts, for example 'research data', 'data lifecycle', 'data preservation', 'data repository', 'data management plan', 'Digital Object Identifier (DOI)', 'metadata', 'Open Access', 'open data', and 'embargoed data';
- Expansion of reasons for RDM, for example making data accessible for re-use in further research, addressing funding bodies and publishers' requirements, and ensuring that data can be used as research outputs;
- Information on other UP policies with which this policy can be associated;
- Clearer stipulation of processes, procedures and responsibilities of role players (for example heads of departments, principle investigators, researchers, promotors, and Departments of Research and Innovation, Library Services, and Information Technology Services) during each of the stages of the RDM process; and
- Information on support and training (Van Wyk, 2013b; Van Wyk, Kleyn & Butler-Adam, 2017).

The policy was reviewed and approved by the University Executive in August 2017. Final approval by the University Senate is expected in the last quarter of 2017 (Pienaar, 2017).

4.8.8 RDM Infrastructure Project

From July 2016 to January 2017, the Departments of Library Services and Information Technology Services collaborated on an RDM infrastructure project to identify a Research Data repository solution for UP (Van Wyk and Van der Walt, 2017). Commercial and Open Source products that could be utilised as a Research Data Repository Platform as part of a total RDM solution for UP, were evaluated (Van Wyk and Van der Walt, 2017). To finalise the product evaluation criteria, various stakeholders were consulted, which included a visit by Library and Information Technology Services staff to peer universities in Australia (Van Wyk and Van der Walt, 2017). Stakeholders were also consulted at a NeDICC workshop on Data Repository evaluation in 2016. Products were short listed through a process of scanning of products that were being used internationally by universities similar to UP in size and research activity (Van Wyk and Van der Walt, 2017). Evaluation criteria included: functional criteria, non-functional criteria, technical aspects, vendor specific criteria, performance requirements, integration requirements and the possibility of consortial pricing (Van Wyk and Van der Walt, 2017). Five products were shortlisted, namely Dspace, Figshare, Islandora (a Fedora-based system), Dataverse, PURR, and Redbox. A Request for Information (RFI) was sent to the vendors/implementation partners of these products. One of the implementation partners failed to respond, while another sent through insufficient information, and one product turned out to cater only for metadata (Van Wyk and Van der Walt, 2017). This narrowed the list to three products: two open source products, namely Dspace and Islandora, and one commercial product, Figshare (Van Wyk and Van der Walt, 2017). Islandora was found to have the best fit, followed by Figshare. Dspace did not have a sufficient fit (Van Wyk and Van der Walt, 2017). Figshare was identified as the preferred option because of a lack of capacity skills, funding, as well as the possibility of consortial pricing (Van Wyk and Van der Walt, 2017). An overview of the process and results of this investigation was subsequently presented at the eResearch Africa 2017 Conference (Van Wyk and Van der Walt, 2017).

As mentioned in 4.4.4.5, various universities across South Africa, including representatives from UP, held an open discussion during the eResearch Africa 2017 Conference with DIRISA and representatives from Figshare, on a possible Roadmap for a South African Figshare consortium (eResearch Africa, 2017). During the subsequent regional meetings arranged by DIRISA in Pretoria and Durban in July 2017, UP also presented, and Figshare was introduced and demonstrated to research institutions and academia (DIRISA, n.d.). UP is partaking in the Figshare trial period, mentioned in 4.4.4.5, by using one of the pilot projects mentioned in 4.8.3 to experiment with Figshare and to provide feedback (Van der Walt, 2017).

4.8.9 Involvement In Data / Library Carpentry

UP's Department of Library Services' first involvement in Data/Library Carpentry came in 2016, when three of its members attended the first Library Carpentry Workshop in Africa. The workshop was hosted by NWU and the company Talarify in collaboration with NeDICC, at the Knowledge Commons of the CSIR in Pretoria, South Africa (NeDICC, 2016). Data Carpentry is a movement that develops and teaches, in workshops, the fundamental data skills that would be needed to conduct research (Data Carpentry, n.d.). The focus is on introductory computational skills such as cleaning data with Open Refine, tools for collaboration such as Git and Github, data management with SQL, and data analysis and visualization with R and Python (Data Carpentry, n.d.). Library Carpentry is not geared towards teaching Library professionals computational skills. Two members of the Department of Library Services attended a Data Carpentry Workshop for instructor training in May 2017 (Van der Walt, 2017).

4.9 SUMMARY

In this chapter, the researcher provided an overview of what is meant by the concepts data and research data, as well as the concepts related to the management of research data, such as data curation, data stewardship, data governance, data archiving, and data management. Thereafter, the researcher gave an overview of a number of international developments with regards to RDM, followed by a comparison of the similarities and differences in these different approaches to RDM. The South African situation on RDM was deliberated upon next, which included a discussion on government initiatives, national collaborative initiatives, initiatives at higher education institutions, other initiatives, and potential partners. This was followed by an exploration of the research data cycle, which included a comparison of a number of cycles from literature. Following this, the researcher covered the different stages of a research data lifecycle as well as the corresponding processes that take place in each, and the potential role that the various stakeholders can play in each. Included in the discussion were processes that take place throughout the whole lifecycle. After this, the concept of big data was addressed. Following this, the researcher discussed the value that RDM has for researchers and institutions. Lastly, the researcher provided an overview on the developments regarding RDM at the University of Pretoria.

In the first two chapters, the researcher discussed the concept of VREs and indicated that RDM is a component of a VRE. This was followed by a discussion on RDM in this chapter. The complexity of managing research data necessitates the use of a VRE to accomplish this. The relationship between RDM and VREs is discussed in the next chapter.

CHAPTER 5
RESEARCH DATA MANAGEMENT AND VIRTUAL RESEARCH ENVIRONMENTS (VREs)

5.1 INTRODUCTION

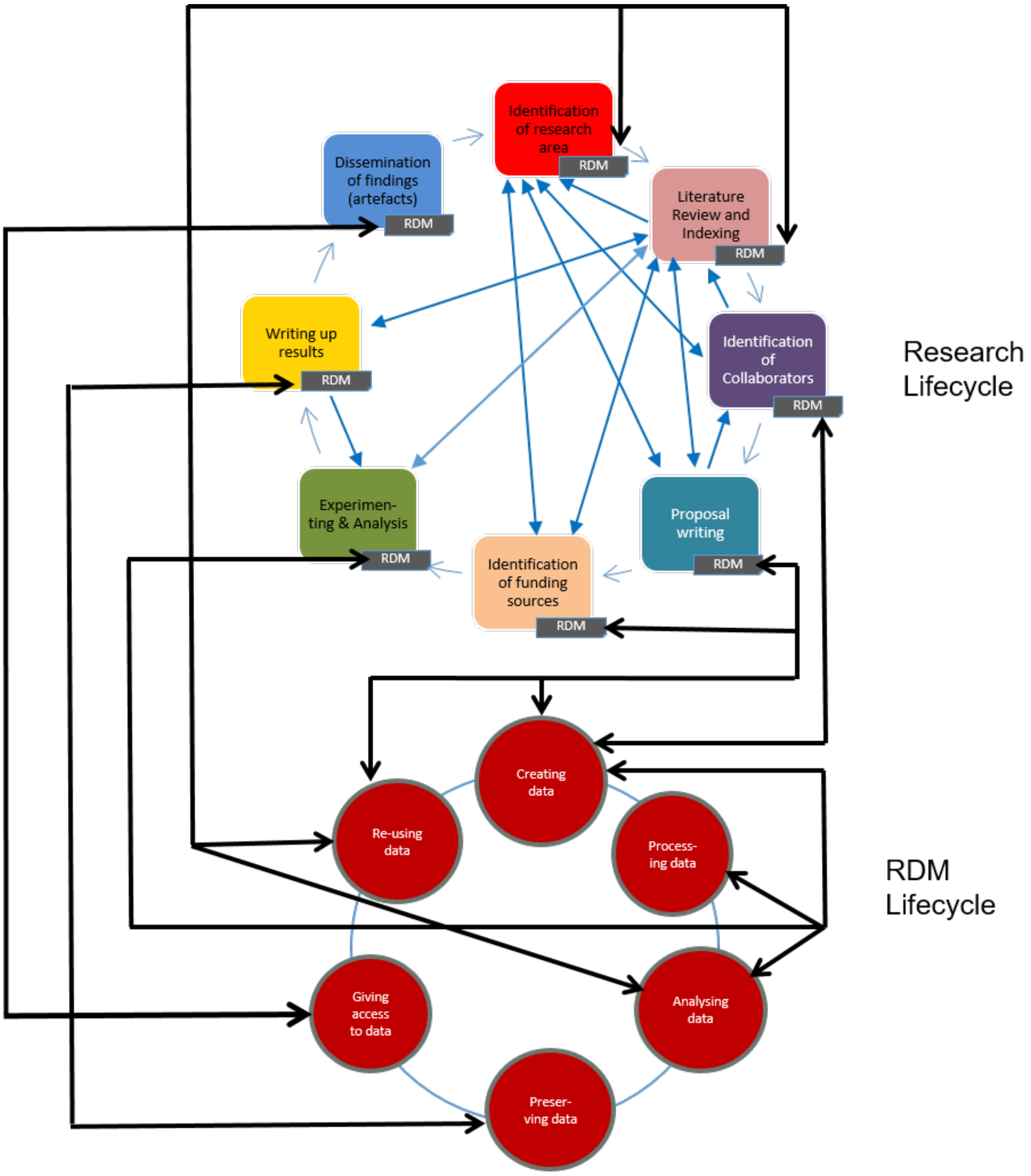
The discussions on the concepts RDM and VREs in the previous chapters revealed that various authors have written about these concepts. Exploring the relationship between these two concepts, however, showed little research done on this area, with many authors, for example Anderson, Dunn and Hughes (2005), Brown (2013), Carusi and Reimer (2010), Filetti and Gnauck (2011), Fraser (2005), Thanos (2013), and Van Deventer et al. (2009), assuming that these terms are intertwined, or that RDM is a given in a VRE. This chapter first aims to explore the research data lifecycle and its relation to the research lifecycle. Following this, the role of RDM as a component within a VRE is examined, followed by a discussion of the management of research data by means of a VRE. Finally, a possible conceptual model for the management of research data by means of a VRE is developed.

The next section will explore the relationship between the research data lifecycle and the research lifecycle.

5.2 RESEARCH DATA LIFECYCLE AND ITS RELATION TO THE RESEARCH LIFECYCLE

The research data lifecycle (See Figure 4.7) as discussed in 4.5.2 does not function independently from the research lifecycle (See Figure 3.4), but is interwoven with the research lifecycle. JISC published a Guide titled *Implementing a Virtual Research Environment* in 2013, which was updated in 2016 (JISC, 2016b). In this guide, they illustrate the inter-relatedness of the research lifecycle and research data lifecycle. Their research lifecycle consists of the following components: Ideas, Partners, Proposal Writing, the Research Process and Publication. They linked the research data management lifecycle to the Research Process stage of their research lifecycle (JISC, 2016b). The research process, however, has a number of stages that were not shown in their illustration, while proposal writing and publication / publishing, in the viewpoint of this researcher, are stages in the research process. The research data lifecycle also impacts on these two stages of the research lifecycle, but in JISC's (2016) model, does not seems to be impacted. The different stages of the research data lifecycle and their relationship to the different stages of the research lifecycle is also not clearly shown. Another shortcoming of this illustration of JISC (2016) is that the iterativeness of the lifecycles are not clearly defined. Figure 5.1 illustrates the interrelatedness of these two cycles, as well as their iterativeness.

Figure 5.1: Integrated Model Of Research Data Lifecycle And Research Lifecycle


Identification of research area
RDM
Dissemination of findings (artefacts)
RDM
Literature Review and Indexing
RDM
Writing up results
RDM
Identification of Collaborators
RDM
Research Lifecycle
Experimen-ting & Analysis
RDM
Proposal writing
RDM
Identification of funding sources
RDM
Creating data
Re-using data
Process-ing data
RDM Lifecycle
Giving access to data
Analysing data
Preser-ving data

The interdependence between the two cycles can best be illustrated by matching the components of the research data lifecycle with the stages of the research lifecycle, as can be seen in Table 5.1.

Table 5.1: Matched Components Of Research Data Lifecycle With Stages In The Research Lifecycle

Research Lifecycle Stages	Components from Research Data Lifecycle	Action
Identification of Research Area	[Re-using Data] [Analysing Data]	• Data Repurposing • Data Discovery
Literature Review and Indexing	Re-using Data [Analysing Data for re-use]	• Search Data Journals • Data Discovery
Identification of Collaborators	Creating Data Re-using Data	• Data Mining
Proposal Writing	Creating Data Re-using Data	• Designing DMPs • Data Repurposing • Data Storage • Data Citation
Identification of Funding Resources	Creating Data (list of potential funders) Re-using Data	• Designing DMPs
Experimenting and Analysis	Creating Data Processing Data Analysing Data	• Data Cleansing • Data Verification • Data Validation • Data Anonymisation • Data Visualisation • Data Interpretation and Analysis • Data Storage • Metadata Creation
Writing up Results	Preserving Data	• Data Archiving/Long-term Preservation • Metadata Creation
Dissemination of Findings	Giving Access to Data	• Data Publishing • Data Sharing (e.g. Social Media) • Add DOI • Data Citation • Link Data to Outputs
Identification of New Research Area	Re-using Data	• Data Repurposing

In Table 5.1, each of the components of the research data cycle have been matched with the corresponding stage of the research cycle. In the ‘Identification of Research Area’, researchers can do an information search to discover resources that are available. This can include data sets (See Referencing Data in 4.2.1) that have been made available by other researchers in repositories, data archives, data centres, or as part of a publication. These data sets can then potentially be re-used to formulate new hypotheses, and generate new cycles of research. In some cases, a researcher can through the analysis of data identify a new area of research (data discovery), and then formulate a new hypothesis (data repurposing). A more recent development has also been the publishing of data articles in data journals, which links up with the ‘Literature Review and Indexing’ stage. These articles discuss data sets and are valuable sources of information. These journals can be searched, and can lead to the discovery of data sets that can be re-used for further research. The data that is found in these articles can also be analysed for further use, which can lead to re-use and discovery.

The ‘Identification of Collaborators’ stage is the stage where the researcher can do a search of literature or data sets (data mining) to establish who the experts in a specific field or topic are (See Referencing Data in 4.2.1). The results of such a process can be the compilation of a list of collaborators (data creation), which can be re-used by other researchers for further research. In the proposal writing stage of the research cycle, the researcher can re-use existing data (repurposing) or can create new data. This is the stage where the researcher draws up a RDM plan, which will present an overview/parameter of the research that he/she plans to undertake. Data generated or used in this early stage of the research cycle should be stored in a secure but accessible place, and should be cited/referenced in a correct manner to make it findable again. The next stage is the identification of funding resources. In this stage, the researcher can create data by drawing up a list of potential funders or funding organisations (See Referencing Data in 4.2.1). These funders or funding organisations will also influence the use and nature of the design of RDM plans. Potential funders could furthermore be identified by re-using data from existing lists of funders.

During the ‘Experimenting and Analysis’ stage, the researchers create, process and analyse the data through processes of cleansing of data, verification of data, validation of data, visualisation of data, and anonymisation of the data, followed by an interpretation and analysis of the data. During this stage, it is also essential that the data generated are stored in a secure place, and that metadata is added to the data to ensure its findability. The ‘Writing up Results’ stage can be matched to the “preserving data” component of the research data lifecycle. Many of the actions taken in the ‘Experimenting and Analysis’ stage can be repeated in the ‘Writing up the Results’ stage, which emphasizes the iterative nature of the research process. The results can then be archived/preserved for a long time. During the ‘Dissemination of Findings’ stage, the researcher can give access to his/her data for re-use. A DOI (a persistent identifier) could be added to each data set. This can then be included when drawing up a citation to the data. In this stage, data could also be linked to the research output that is based on that data; for example, a research article. This is the stage where the researcher publishes his/her data on a repository, or data archive, or data centre, or as part of a research article. Data is sometimes also shared with other researchers via social media or e-mail. The sharing of data then enables others to identify ‘New Research Areas’ by re-using the data and repurposing it for new research. By doing so, it generates a new research cycle.

5.3 A VRE AS AN ESSENTIAL FRAMEWORK FOR THE MANAGEMENT OF RESEARCH DATA

The management of research data typically functions within an infrastructure that enables researchers to access, manage and utilise data. VREs can provide such an infrastructure. This is confirmed by various authors in their definitions of VREs. Carusi and Reimer (2010: 13), for example, see a VRE as facilitating collaboration between researchers and providing access to data, tools and services through a technological framework that accesses a wider research infrastructure. The DFG describes a VRE as an internet-based collaborative working platform that facilitates a new method of “dealing with research data and information” (Carusi and Reimer, 2010: 14). Thanos (2013: 77), as mentioned in 2.2.8.1, views it “as a framework within which data tools and services can be plugged.” In addition, Van Deventer et al. (2009) list data production, data retrieval, data analysis, and data visualisation as some of the processes VREs aim to support. This is confirmed by Filetti and Gnauck (2011: 238) when they list data analysis, data visualisation, and data warehousing (to provide complex data storage as well as data analysis), as possible services of a VRE.

These definitions show that RDM is indeed an important component within a VRE, but how is research data managed within a VRE, and what role does it play in a VRE? These questions are answered in the next section.

5.3.1 Managing Research Data By Means Of A VRE

A number of the characteristics of VREs listed in 2.2.8.3 emphasize the usage of VREs in managing research data. Brown (2013) and Robertson Library (n.d.) indicate that a key characteristic of a VRE is that it affords researchers and research teams more effective ways as well as the tools necessary for collecting (capturing), manipulating, managing and securing data collaboratively. Another characteristic as mentioned by Carusi and Reimer (2010: 19) shows that VREs could be used for analysis and processing of data, annotating data collaboratively, and sharing of data with peers. This sharing of data aspect is emphasized by Filetti and Gnauck (2011: 237) as a key element in a VRE. The interdisciplinary nature of VREs also allows for the gathering of data and approaches from different disciplines to create new research findings (Carusi and Reimer, 2010: 23; Fraser 2005). A VRE can further provide researchers with new forms of data and challenges to analysis (Wilson et al., 2007: 290).

Carusi and Reimer's (2010: 18-19) VRE Collaborative Landscape Study also showed that integrating an architecture for data management into a VRE can address the issue of preservation of research data, as it can provide an easy to use technological framework where researchers can secure the short-term storage of their data, and also afford them the means to keep control of their work. This is confirmed by Neuroth, Lohmeyer and Smith (2011: 225) in their discussion on the TextGrid VRE. According to them, the advantage of "combining the tools and services for text-based research" in a VRE with a data management system, will provide a safe place to researchers (in this case, grid storage) where they can save their data directly (Neuroth, Lohmeyer and Smith, 2011: 225). Uploading research data directly onto an institutional repository, without using a VRE, is possible nevertheless, but these "repositories can often seem somewhat alien to them", as they are not normally structured "in the way that the researchers work" (it is not part of their research workflow) (Carusi and Reimer, 2010: 18). Researchers will however be encouraged to use a VRE if it arrives with "a well thought out data management plan and the tools" necessary "to use and create data in documented formats" (Carusi and Reimer, 2010: 18). Most VREs are also fully integrated with the research process (cycle), hence providing excellent access points for institutional repositories (Carusi and Reimer, 2010: 18; Neuroth, Lohmeyer and Smith, 2011: 223, 230). The collaborative nature of VREs furthermore provides researchers with the possibility to "share data and collaborate" in collecting, manipulating, analysing and interpreting data (JISC, 2006; Carusi and Reimer, 2010: 20). This is especially true for data as well as for co-researchers that are geographically separated, for example in archaeology. On top of that, VREs provide easy access to computational resources and collaborators, which results in "faster research results and novel research directions" (Carusi and Reimer, 2010: 5; Pham et al., 2005: 16).

Carusi and Reimer (2010: 22) also emphasize the transformative role that VREs can play to take research to the next level; for example, bringing researchers that are geographically dispersed together by providing the necessary tools that will enable them to work more intensively on a project than would have been possible in a once-a-year meeting. This is also mentioned by Anderson, Dunn and Hughes (2005: 516), who give an example of the Silchester VRE that was formed around an archaeological project focusing on the excavation of a Roman Town in Silchester, UK. This VRE had an online conferencing facility, which geographically dispersed researchers could access to mine an integrated database. It also enabled them to have "real-time meetings in the presence of the data" (Anderson, Dunn and Hughes, 2005: 516). Another example is the possibility provided by VREs to integrate articles, comments and data (Carusi and Reimer, 2010, 22; Anderson, Dunn and Hughes, 2005: 516).

A VRE can also be used for access to and location of data (Yang and Allan, 2010: 68). Context and provenance of data furthermore plays a very important role in ensuring that data are trustworthy, and VRE's can provide the necessary rich context to ensure this (Carusi and Reimer, 2010: 42). In addition, research data generated through "models/simulations, observations, and experiments are intrinsically linked with the data collection methodologies and instrumentation" according to Martinez-Uribe and MacDonald (2009: 311). A VRE is the ideal place to position it. Finally, a successful VRE, according to Filetti and Gnauck (2011: 237), will have, among other things, clarity on the ownership of the data and an approved project plan with data policies for the benefit of the collaborating researchers.

5.4 A POSSIBLE CONCEPTUAL MODEL FOR RDM IN A VRE

The conceptual framework of a VRE as illustrated in Figures 3.12(a-c) was adapted to include a more comprehensive overview of the essential component of RDM. This adjusted conceptual framework also consist of a human layer with human components, a hardware layer with possible hardware components, and a software layer, comprising software components. These three layers together with available RDM components and other pluggable VRE components support and impact the research process as well as the RDM process as it develops through the research cycle. Figure 5.2a presents a high-level overview of the framework.

Figure 5.2a: High Level Overview Of Model

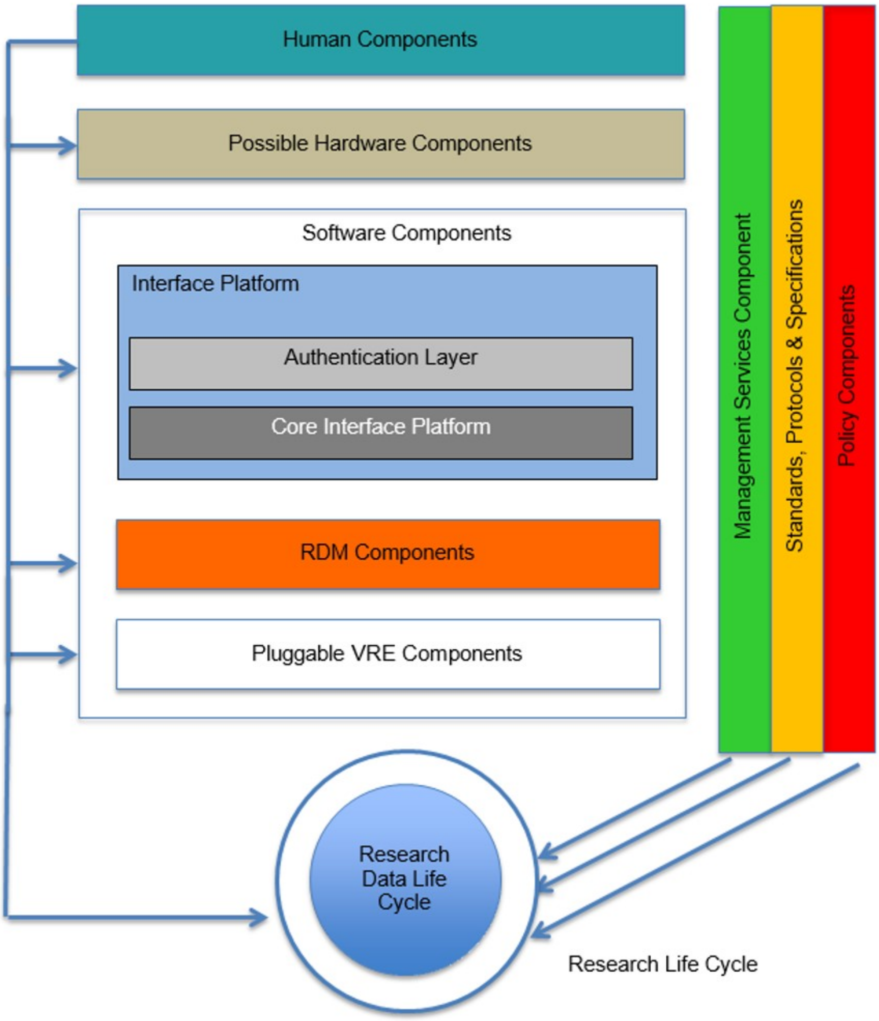


Figure 5.2b: Conceptual Model Showing RDM Components

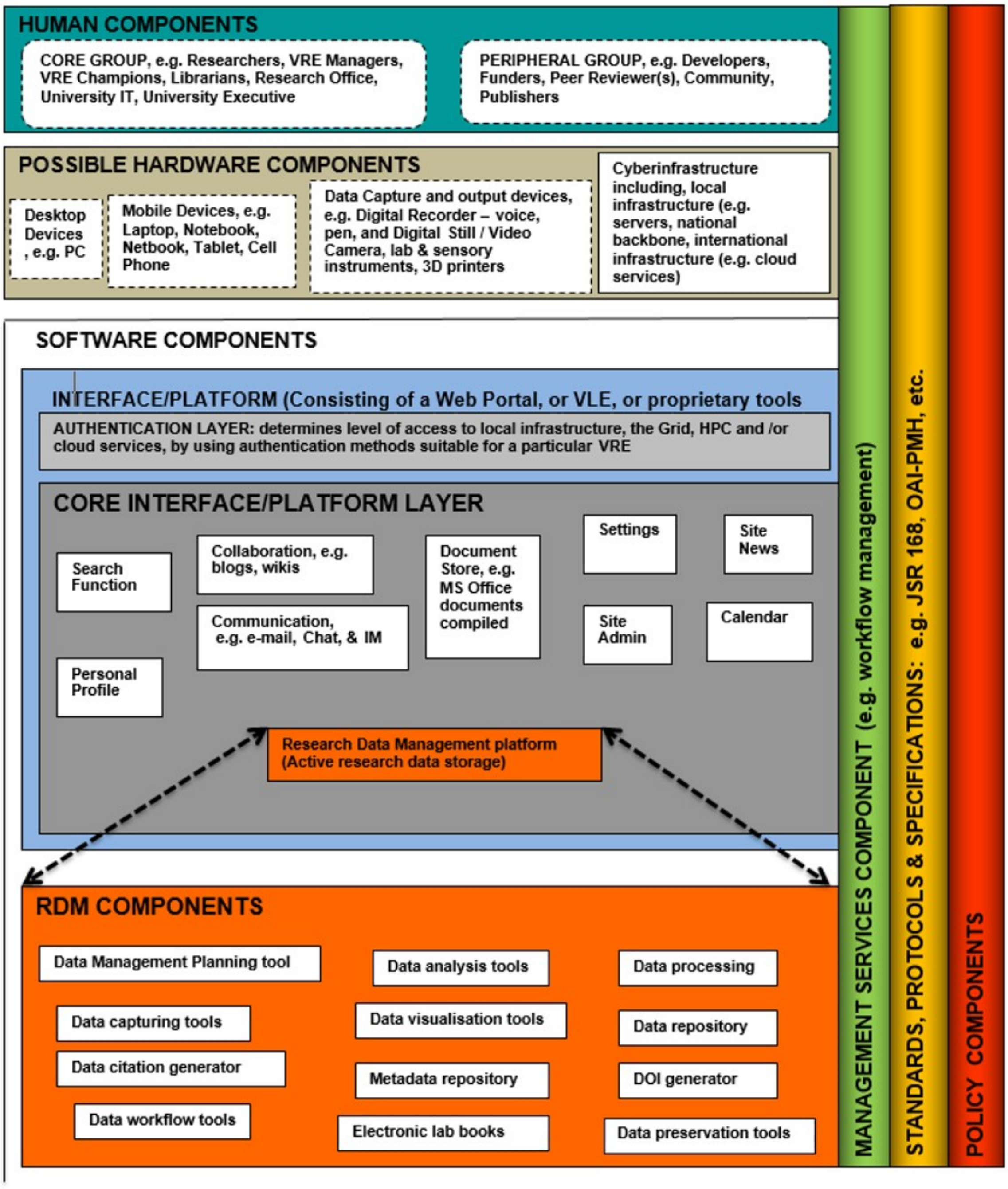
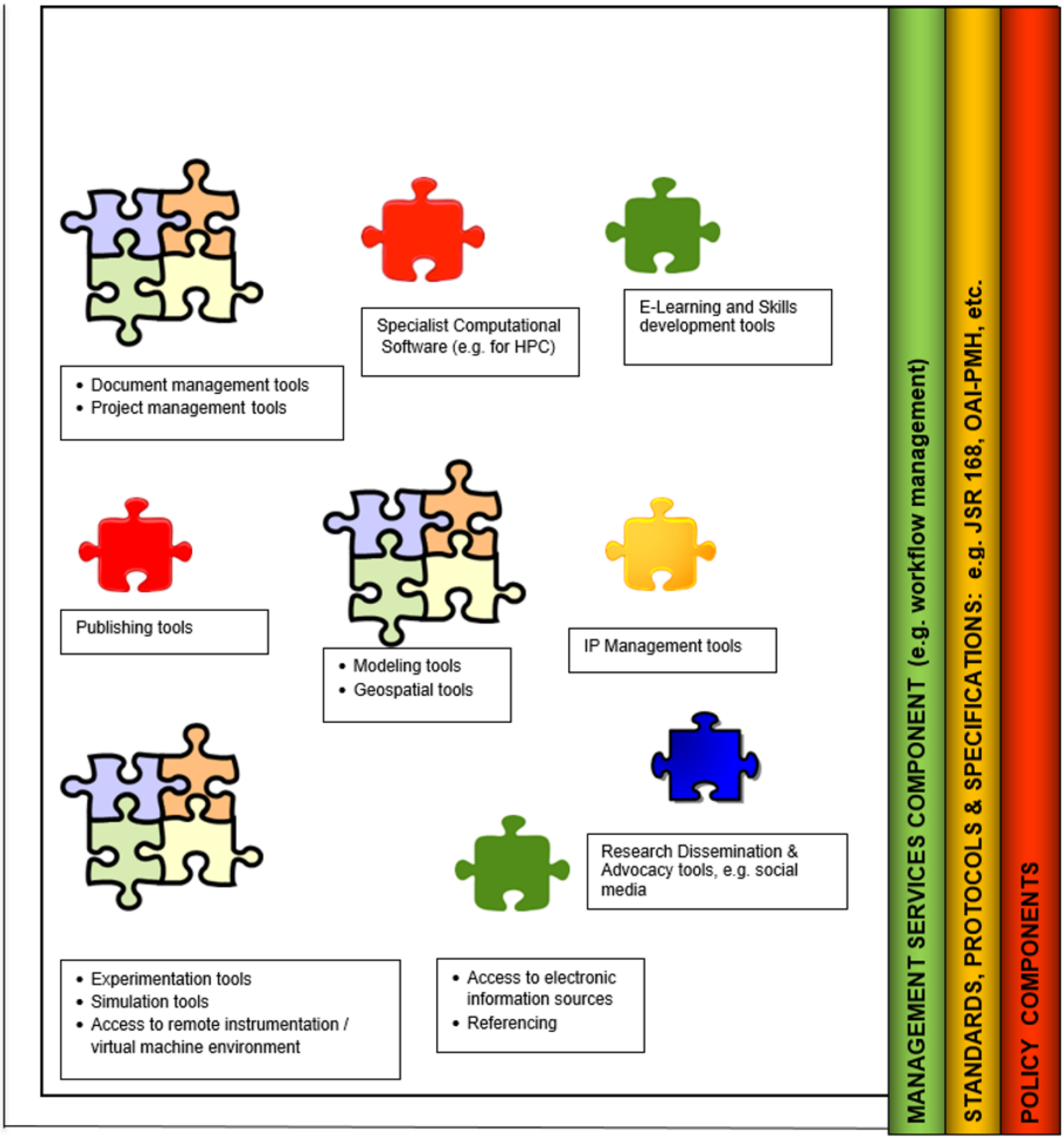

HUMAN COMPONENTS
CORE GROUP, e.g. Researchers, VRE Managers, VRE Champions, Librarians, Research Office, University IT, University Executive
PERIPHERAL GROUP, e.g. Developers, Funders, Peer Reviewer(s), Community, Publishers
POSSIBLE HARDWARE COMPONENTS
Desktop Devices , e.g. PC
Mobile Devices, e.g. Laptop, Notebook, Netbook, Tablet, Cell Phone
Data Capture and output devices, e.g. Digital Recorder – voice, pen, and Digital Still / Video Camera, lab & sensory instruments, 3D printers
Cyberinfrastructure including, local infrastructure (e.g. servers, national backbone, international infrastructure (e.g. cloud services)
SOFTWARE COMPONENTS
INTERFACE/PLATFORM (Consisting of a Web Portal, or VLE, or proprietary tools
AUTHENTICATION LAYER: determines level of access to local infrastructure, the Grid, HPC and /or cloud services, by using authentication methods suitable for a particular VRE
CORE INTERFACE/PLATFORM LAYER
Search Function
Collaboration, e.g. blogs, wikis
Communication, e.g. e-mail, Chat, & IM
Document Store, e.g. MS Office documents compiled
Settings
Site News
Site Admin
Calendar
Personal Profile
Research Data Management platform (Active research data storage)
RDM COMPONENTS
Data Management Planning tool
Data analysis tools
Data processing
Data capturing tools
Data visualisation tools
Data repository
Data citation generator
Metadata repository
DOI generator
Data workflow tools
Electronic lab books
Data preservation tools
MANAGEMENT SERVICES COMPONENT (e.g. workflow management)
STANDARDS, PROTOCOLS & SPECIFICATIONS: e.g. JSR 168, OAI-PMH, etc.
POLICY COMPONENTS

Figure 5.2c: Other Pluggable VRE Components



Document management tools
Project management tools
Specialist Computational Software (e.g. for HPC)
E-Learning and Skills development tools
Publishing tools
Modeling tools
Geospatial tools
IP Management tools
Research Dissemination & Advocacy tools, e.g. social media
Experimentation tools
Simulation tools
Access to remote instrumentation / virtual machine environment
Access to electronic information sources
Referencing
MANAGEMENT SERVICES COMPONENT (e.g. workflow management)
STANDARDS, PROTOCOLS & SPECIFICATIONS: e.g. JSR 168, OAI-PMH, etc.
POLICY COMPONENTS

Figure 5.2d: Model Applied To Research Lifecycle And Research Data Lifecycle

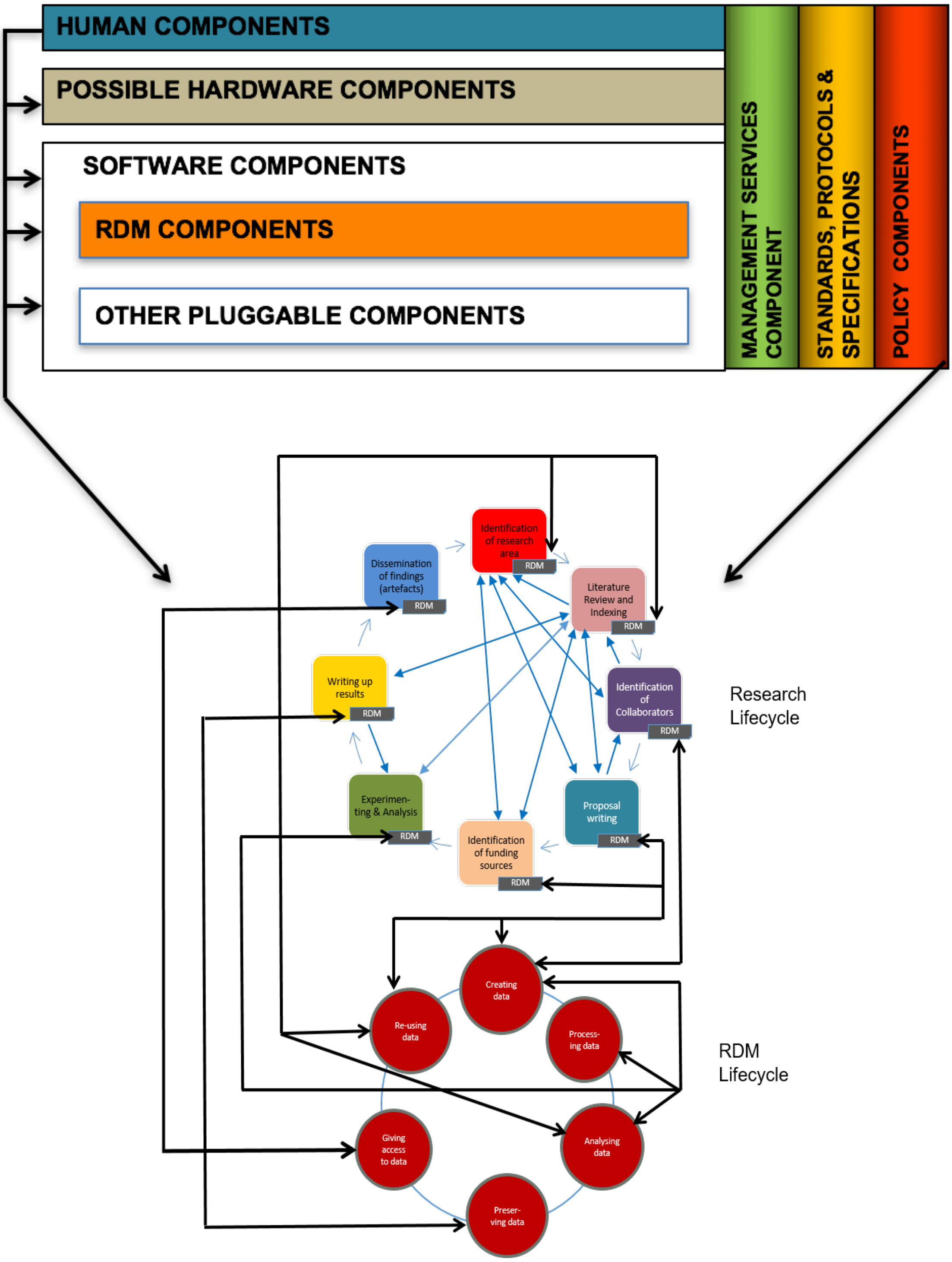

HUMAN COMPONENTS
POSSIBLE HARDWARE COMPONENTS
SOFTWARE COMPONENTS
RDM COMPONENTS
OTHER PLUGGABLE COMPONENTS
MANAGEMENT SERVICES COMPONENT
STANDARDS, PROTOCOLS & SPECIFICATIONS
POLICY COMPONENTS
Identification of research area
RDM
Literature Review and Indexing
RDM
Identification of Collaborators
RDM
Proposal writing
RDM
Identification of funding sources
RDM
Experimenting & Analysis
RDM
Writing up results
RDM
Dissemination of findings (artefacts)
RDM
Research Lifecycle
Creating data
Process-ing data
Analysing data
Preser-ving data
Giving access to data
Re-using data
RDM Lifecycle

Figure 5.2e: Policy Components

- Clear ground rules, e.g. Determine who act as facilitator; Determine the roles in the VRE
- Trust relationships
- Clearly defined objectives
- Mutually agreed project plan/collaborative agreement
- Encouragement of shared interest and enthusiasm
- Intellectual Property (IP) issues across country borders should be dealt with beforehand
- Protection of rights
- Ethical issues must be considered and taken care of
- Proper matching of skills levels and research interests
- Decision on type of interface, type of grid service, and/ or cloud service, pluggable components, standards and protocols
- Negotiations / Decisions on shared access to publications, conference papers (licensing issues)
- Negotiations / Decisions on shared access to research equipment, instruments, and technology
- Negotiations / Decisions on shared opportunities for publishing and presentations
- Regular progress monitoring

5.4.1 The Human Components Layer

This layer consists of the various human actors that might possibly want access to such a VRE. For a more detailed discussion on the various human actors, see 3.5.7.1.

5.4.2 Hardware Components Layer

As mentioned in 3.5.7.2, this layer consists of the various hardware components that can potentially be chosen by the human components in a VRE configuration, or to access a VRE.

5.4.3 Software Components Layer

As discussed in 3.5.7.3, the software components layer itself consists of an interface or platform layer, and a layer with components (applications and services), which can be plugged into the VRE as needed. In the adjusted conceptual framework (Figure 5.2b), the researcher expanded a set of RDM components to emphasize the importance of this for a VRE.

5.4.3.1 Interface Or Platform Layer

In the adjusted conceptual framework, the interface or platform layer still forms the front-end of the software component layer of a VRE, which is seen and accessed by the human components. The authentication layer as part of the interface layer is a critical component to determining levels of access to various parts of the VRE, especially data. Levels of access rights are essential when working with sensitive data or data with commercial value. In 3.5.7.3, the researcher indicated that the various human components would have different access rights that are determined through a registration process and logins and passwords. For a list of potential authentication methods, see 3.5.7.3.

5.4.3.2 Core Interface / Software Layer

In 3.5.7.3, the researcher indicated that the core interface / software layer consists of fixed components that are part of the standard configuration of the specific tool used. These components could vary, but are normally things such as a search function; a personal profile; collaborative writing tools such as blogs and wikis; communication tools such as instant messaging, chat, and e-mail; a document store where documents can be compiled with a word processing system, and stored; a settings function; a site news function; a site admin function; a calendar; and a RDM platform, e.g. a research data store. The RDM platform's functionality could be enhanced, as shown in the adapted conceptual framework (Figure 5.2b), by adding more RDM components; however, at times, some of these components could form part of the RDM platform's core function.

5.4.3.3 RDM Components

A number of RDM components could be added to the VRE. These are:

- A DMP tool, e.g. DMPTool developed by the University of California Curation Center of the California Digital Library, and DMPonline tool developed by the Digital Curation Centre, UK (see 4.5.2.1 for more information);
- A data citation generator, e.g. Elsevier's DataLink (see discussion on data citation in 4.5.2.6.);

- Data capturing tools, e.g. surveys (for example Survey Monkey, and Qualtrics), sensor instruments, experiments using computerised laboratory instruments, etc. (see discussion on data capturing in 4.5.2.1);
- Data analysis tools, e.g. R, SAS, SPSS, Qualtrics (see discussion in 4.5.2.3);
- Data processing tools;
- Data workflow tools, e.g. Taverna, which captures the automated repetitive scientific processes that are used to generate data (Harvey, 2010: 49);
- Data visualisation tools, Excel graphs and figures, Google Maps, Visual.ly, Tableau, ArcGIS, etc. (see discussion in 4.5.2.3);
- Electronic lab books, e.g. Labarchives, Accelrys, etc;
- Data repository (publishing), using software such as Fedora, or DSpace, Eprints, FigShare, CKAN, etc. (see discussion on data publishing in 4.5.2.3);
- Digital Object Identifier (DOI) generator (see discussion on DOIs in 4.5.2.6 under Data citation);
- Metadata repository, e.g. CKAN, Repository in a Box, etc. (see discussion on metadata in 4.5.2.1); and
- Data preservation tools, e.g. the Bagger application, which was designed for the Library of Congress as a tool to produce a package of data files according to the BagIt specification, Archivematica, etc. (see discussion on data preservation in 4.5.2.4).

5.4.3.4 The Bottom Layer Of The Software Components Layer

The bottom layer of the software components layer (Figure 5.2c) comprise various other software components (services and applications) that can be ***plugged*** into the interface/platform component determined by the needs of each VRE community/project. As discussed in 3.5.7.3, these components can vary, but the following components might be needed (grouped together by related function):

- Document management tools, and project management tools;
- Specialist computational software (for example to use for HPC, and sequencing);
- E-learning and skills development tools;
- Publishing tools;
- Modelling tools and geospatial tools;

- Intellectual property management tools;
- Access to electronic information sources and referencing tools, e.g. Endnote;
- Experimentation tools, simulation tools, and access to remote instrumentations / virtual machine environments; and
- Research dissemination and advocacy tools, e.g. social media.

5.4.4 Management Services Component (Vertical Layer In Green)

The management services component as discussed in 3.5.7.4 confers automatic behaviour to the whole VRE across the different layers and components by utilising standards, protocols and specifications in service invocation, and has been kept without any adjustments in Figure 5.2b. For more detail see 3.5.7.4.

5.4.5 Standards, Protocols And Specifications (Vertical Layer In Amber)

The various sub-layers within the software components layer are held together by interoperable ***standards, protocols and specifications***. This has not been changed in Figure 5.2b. A more detailed discussion on these standards, protocols and specifications can be found at 3.5.7.5.

5.4.6 Research Lifecycle And Research Data Lifecycle (Figure 5.2c)

The aim of a VRE (with all of its components, including human, hardware, software, standards, protocols, specifications, management services, and policy components) is to support and enhance the research cycle and integrated with that, the research data cycle, each with their distinct components (See Figure 5.2d). The research cycle chosen for this study comprise of the researcher's adapted version of Pienaar and Van Deventer (2009) and Van Deventer et al.'s (2009) research cycle, consisting of the following components, which function iteratively and not necessarily in a cycle: identification of research area; literature review and indexing; identification of collaborators; proposal writing; identification of funding sources; experimentation and analysis; RDM; writing up results; and dissemination/output of findings. The research data lifecycle includes the following components, which can be applied throughout the research lifecycle: creating data, processing data, analysing data, preserving data, giving access to data, and re-using data.

5.4.7 Policy Components (Vertical Layer In Red)

Figure 5.2e provides an expanded view of the policy components of a VRE. These components are essential to ensure the successful operation of a VRE. For a more detailed discussion of these components, see 3.5.7.6.

5.5 SHORT OVERVIEW / SYNOPSIS OF THE LITERATURE REVIEW

In the first literature review (Chapter 2) the researcher of this study first framed the concept of VREs within the wider concepts of e-Science and e-Research and indicated that this study followed the Social Science Approach to e-Research, including the Computerisation Movement, Information Systems, Service-Oriented Architecture, and Whole Process approaches. Various related concepts to VREs were discussed next, namely Cyberinfrastructure, Science Gateways, Cyberscience, Web-based Research Support Systems (WRSS) and Collaboratories. The differences between these concepts and their relationships to VREs are illustrated in Table 5.2.

Following this, the researcher gave an overview of the concept of ‘Virtual Research Environments’, which included a discussion on its definition, development, aims, and characteristics. A definition for a VRE was then formulated by the researcher, namely:

> A VRE consists of a common, flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts etc.) interoperating with each other, can be plugged, to enable collaboration and to support and enhance large and small scale processes of research, which are often performed by researchers in multidisciplinary contexts, within or across organisational and geographical boundaries.

Table 5.2: Related concepts to VREs

	Cyberinfrastructure	Science Gateways	Cyberscience	WRSS	Collaboratories
Function	This is infrastructure, based upon distributed computer and information technologies.	These are community-developed bundles of tools, applications and data collections for a targeted community interface to cyberinfrastructure.	This covers all scholarly and scientific research activities in the virtual space, generated by networked computers and information communication technology.	These systems develop new and effective software systems and tools for research institutions and researchers to support their research activities.	These are Web-based collaboration environments where researchers can access instrumentations, computational resources and data.
Relationship with VREs	VREs operate over cyber-infrastructure.	This is similar to VREs, but is a term mostly used in the USA.	This is a term similar to e-Research, but is not used anymore. It is therefore a broader term than VREs.	This term is synonymous with Web-based VREs.	This term is used in the USA and the Netherlands, but has been supplanted by the VRE concept.

In the second literature review (Chapter 3), the researcher gave an overview of the development of VREs internationally by focusing specifically on four countries: the UK, the USA, the Netherlands and Germany. These countries were chosen as they are representative of different VRE approaches or models used across the globe. Similarities and differences between the approaches in these countries were discussed under five groupings: organisational aspects, technical aspects, functional aspects, policy/legal/financial aspects, and cultural aspects.

The discussion under organisational aspects showed that the UK and the USA had a bottom-up and user-driven approach, which included users in the design process and in the technology and software used. In all four countries, users were given the freedom to experiment and develop their own technologies, or to adapt existing ones. In Germany, funded projects were required to be collaborations between researchers and developing institutions such as libraries, computer centres or e-research centres.

The discussion under technical aspects revealed that in the UK, some projects were shelf-ready projects, while others used content management tools, portal technologies, or general institutional web-based tools. In Germany, the development of new software following open source principles were encouraged and funded, as were existing solutions. In the USA, the focus was more on the development of science gateways, portal technology, and gridware, as well as the creation of hubs (cloud-driven tools). The Netherlands had a flexible approach that gave funded projects the freedom to test any environment.

The discussion under functional aspects disclosed that all four countries focused on collaboration and sharing of ideas, and on supporting the research lifecycle. All four also supported single, interdisciplinary and cross-institutional research. Data sharing was revealed to be an important and central feature in both the Netherlands and Germany, while the Science Gateways project in the USA showed that that there is an increasing interest among researchers to make more data available. The VRE programme in the UK further revealed the importance of evaluating and assessing the success of VRE projects. The focus on VREs in the UK and the Netherlands also led to rethinking of research methods in projects, and the Science Gateways project in the USA was found to provide useful interfaces to supercomputing resources.

The discussion under policy/legal/financial aspects showed that in the UK, the Netherlands, and Germany, a national institution funded and drove the main VRE initiatives, whilst in the USA, the NSF funded TeraGrid and HUBzero. The UK also had a joint government-commercial venture between the British Library and Microsoft, to develop RIC. In all four countries, the major challenge was shown to be the sustainability of VREs. The problem of sharing resources that required institutional subscriptions was highlighted by the Netherlands, while the MyExperiment project in the UK revealed that total open access to everything was not possible. Germany stressed the importance of librarians in checking that open access to publications do not violate third-party rights.

The discussion under cultural aspects revealed that the UK was well-advanced in its understanding of the VRE concepts. The Netherlands were shown to be more focused on the humanities and social sciences, whereas in other countries, the focus was more mixed. Feedback from projects in Germany revealed the difficulty in building appropriate services and solutions across discipline boundaries, and trust was stressed as the key factor in the uptake of VREs in Germany.

Following this discussion, the researcher focused on the concept of research cycles, and used Pienaar and Van Deventer's (2009), Van Deventer et al.'s (2009), and Van Deventer's (2015) research cycle as the basis for an adapted research lifecycle in Figure 3.4, which illustrates the iterativeness of such a cycle clearly. This cycle was shown to include the following stages: Identification of a research area; literature review; proposal writing, identification of funding resources, experimenting and analysis, writing up of results, and dissemination/output of findings/closure/continuation.

After this, the researcher discussed various VRE components as proposed by authors such as Myhill, Shoebridge and Snook (2009); Voss and Procter (2009); Chambers (2002); Klyne (2006); Sergeant, Andrews and Farquhar (2006); Di Muro and Saunders (2008); Keraminiyage, Amaratunga, and Haigh (2009); and Van Till and Dovey (2010). These were combined and matched to the different stages in a research lifecycle, where they could have the greatest impact or provide support, together with potential tools (See Table 3.2).

A review of literature (De Roure et al., 2009; Fernihough, 2011: 101; Keraminiyage et al. 2009b: 129-142; McLennan and Kennell, 2010; Simeoni et al., 2008; Yang and Allan's, 2007) disclosed various kinds of VRE models, but they all fell short of providing a possible complete conceptual framework for a VRE. Each of these models, illustrations or frameworks, however, contributed valuable components that could be included in a possible comprehensive conceptual framework. The researcher then designed a possible conceptual VRE model (See Figures 3.12a-c) by combining some of the valuable components provided in the different models, frameworks and illustrations, to get a clearer understanding of how these components interact in a research cycle. This first version of the conceptual VRE model was set up in Figures 3-12a-c and then further developed and expanded in Figures 5.2a-e. This model consisted of a human components layer (with a core group and peripheral group of human components); a hardware components layer with desktop services, mobile devices, data capture and output devices, and Cyberinfrastructure; a software components layer comprising an interface or platform layer, and a core interface layer that contains fixed components that are part of the standard configuration of the specific tool used, as well as a software components layer that can be used or plugged into the core interface or platform. This software components layer was expanded and divided into an RDM components layer and a Pluggable VRE components layer in Figures 5.2b and 5.2c. At the right side of the model there are a number of vertical components layers. The first of these layers is a management services components layer that confers automatic behaviour to the entire VRE across the different layers and components by utilising standards, protocols and specifications in service invocation. Next to that is the standards, protocols and specifications layer, which holds the various sub-layers within the software components layer together. Next to that is the policy components layer containing policy components that are essential for the successful operation of a VRE. All these layers and their components support and enhance the research cycle and integrated with that, the research data cycle.

The third literature review chapter (Chapter 4) focused on RDM, and to gain a better understanding of the process of RDM, the researcher defined the concepts of data and research data and identified the various types of data that could potentially be found in a VRE. Next, the researcher explored various concepts that describe the management of research data. These concepts are: data curation, data stewardship, data governance, data archiving, and data management. The focus of each of these concepts, each one’s focus, and their relationship to RDM, is illustrated in Table 5.3.

The above was then followed by an overview of RDM. The researcher also formulated the following definition of RDM:

> RDM is the process of controlling and organising the data generated during a research project, and covers the entire data lifecycle, which includes the planning of the investigation, conducting the investigation, storage and backing up of the data as it is created, preserving the data long-term, after the research investigation has concluded, and making the data accessible for future use.

Table 5.3: Related concepts to RDM

	Data Curation	Data Stewardship	Data Governance	Data Archiving	Data Management
Function	Operational	Tactical	Strategic	Operational	Strategic, Tactical and Operational.
Focus	Promotes the use of data from its point of creation.	Takes responsibility for datasets.	Focuses on the people managing the data.	Focuses on the storage and collection of data into archive collections.	Focuses on the management of the full data lifecycle needs of an organisation.
Relationship to RDM	Subset of RDM	Subset of RDM	Subset of RDM	Subset of RDM	Not a subset of RDM.

After this, the researcher provided an overview of a number of international developments on RDM. This discussion did not to give an exhaustive overview, but focused on the most important country developments as reflected in literature available in English, and showed the similarities and differences in the approaches to RDM in these countries. The countries / regions that were covered, included the UK, EU, USA and Australia. The discussion on similarities and differences revealed that the UK had the earliest development of RDM services. In the USA, RDM initiatives were mostly driven by mandates received from various funders, while in the UK, these were driven by funding received from government, as well as by mandates from funders. RDM initiatives in the EU and Australia were mostly driven by funding and development of infrastructure for RDM that were provided by government. The literature review also revealed that the UK and USA’s RDM initiatives contributed valuable tools to assist in the RDM process, e.g. DMPOnline and DMPTool. The UK and Australia, in addition, have very useful examples and materials available on developing RDM policies. The literature review also disclosed that libraries in all four countries were actively involved in developing RDM services for their institutions.

To give background and context to the case studies, the South African situation on RDM was deliberated upon next, with a discussion on government initiatives, national collaborative initiatives, initiatives at higher education institutions, other initiatives and potential partners. Following this, the researcher explored the concept of a research data cycle by comparing a number of cycles from literature. The different stages of a research cycle as well as the corresponding processes that take place in each, and the potential role that the various stakeholders can play in each, was identified next. The concept of big data was also touched on. This was followed by an investigation into the value that RDM provides, and an overview on the developments regarding RDM at the University of Pretoria, South Africa - the location of the case studies this study focuses on.

Chapter 5 explored the relationship between RDM and VREs and included a discussion on the relationship between the research lifecycle and the research data lifecycle, and a discussion on the management of research data by means of a VRE, followed by a proposed conceptual VRE model with all its components layers and components, which was developed from information that was gained from the literature review. This model is later tested in the empirical part of this study.

5.6 SUMMARY

This chapter showed that there is relationship between the concepts RDM and VREs. The relationship of the research data lifecycle to the research lifecycle was discussed first and illustrated in Figure 5.1. This was followed by a discussion highlighting the role of RDM as a component within a VRE, as well as a discussion on how a VRE can be utilised in managing research data. Following this, the researcher proposed a possible conceptual VRE model, which was refined from the model that was proposed in Figures 3-12a-c. This refined model was developed through information gained from the literature review in chapters 2-5, and clearly illustrates the management of research data by means of a VRE. The chapter concluded with a short overview / synopsis of the literature review.

The next chapter will give an overview of the research methodology used for this study.

CHAPTER 6
RESEARCH METHODOLOGY

6.1 INTRODUCTION

The literature study in chapters 2-4 commenced (in Chapter 2) with a discussion of the concept of VREs as part of e-Research infrastructure, and their relationship to other concepts such as e-Science, cyberinfrastructure, science gateways, cyberscience, e-Research, collaboratories, and WRSS. This was followed by a discussion in Chapter 3 of the components and tools used in VREs, which included an overview of the development of VREs across the world, and concluded with a proposed conceptual framework of a VRE. In Chapter 4, the concept of RDM, its development internationally and nationally, as well as the various actions that could be taken within the research data cycle, were discussed. The aim of Chapter 5 was to investigate the relationship between VREs and RDM as a component of a VRE. The results found in the literature study (Chapters 2-4) were then applied in practice by focusing on two case studies. This chapter provides an overview of the research design followed. It gives an overview of the non-empirical part of this study, which consists of a literature study of the concepts, and the empirical part of the study, consisting of case studies. The chapter describes what is meant by a literature study and then proceeds to describe what is meant by a 'case study' method. This is followed by a description of the various methods used in the case study, namely sampling method and triangulation, the PAR method, and prototyping. The discussion then focuses on the various data collection methods used, namely participant observation, interviews, as well as testing and prototyping. An overview is subsequently given of the research questions asked during the interviews, followed by a description of the methods of analysis and evaluation.

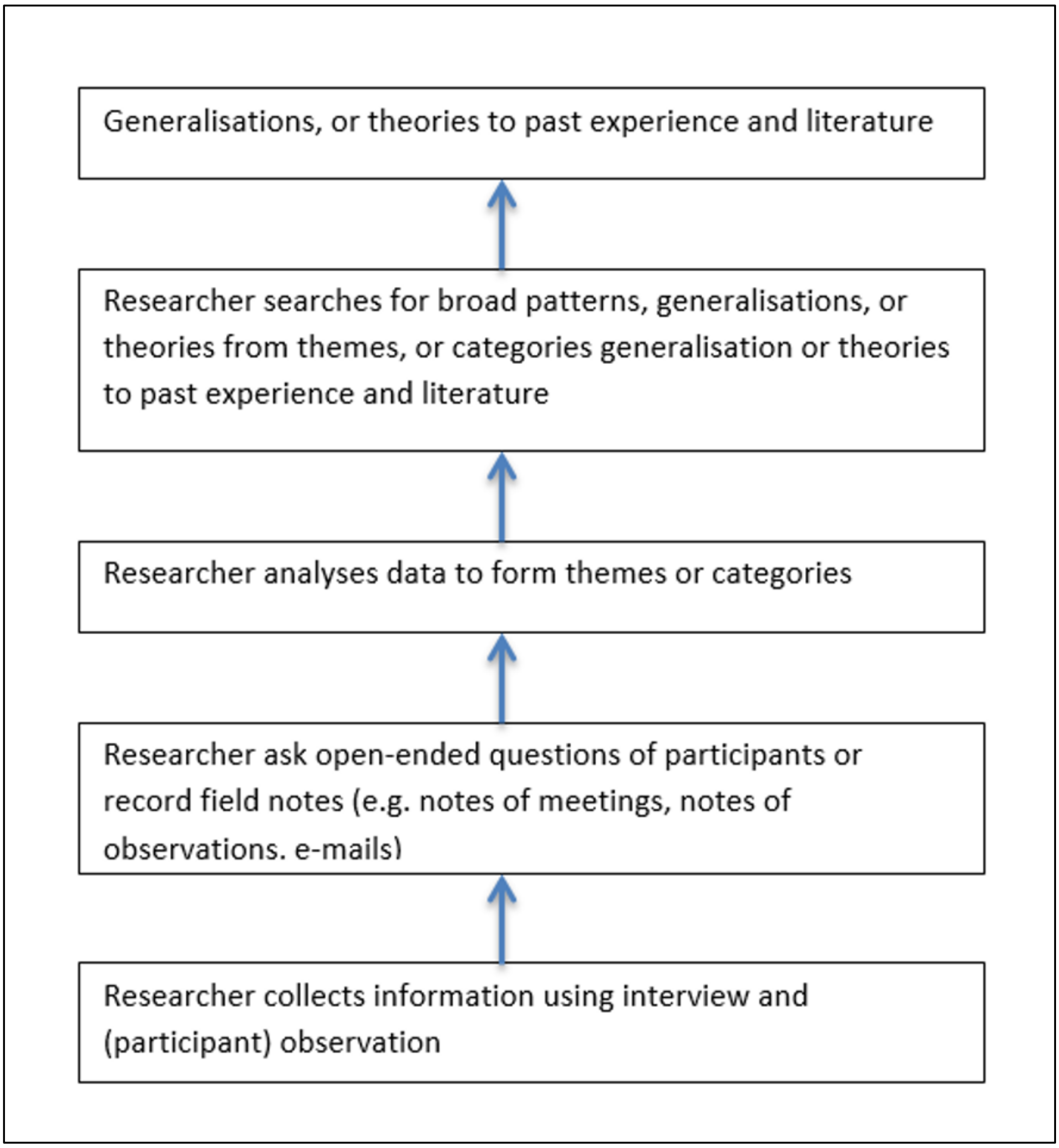
6.2 RESEARCH DESIGN

Mouton (2001: 55-56) describes research design as "a plan or blueprint of how you intend conducting the research." According to this description, one starts with an end product in mind, then formulates a research problem or question, followed by focusing on the logic of the research (in other words, what kind of evidence is needed to address the research question sufficiently). The research design of this study followed an interpretivist paradigm, and specifically, empirical interpretivism. Empirical interpretivism, according to Pickard (2007: 11), focuses on the investigation of social phenomena in natural settings. In this paradigm, knowledge is intentionally acquired through interpretation and meaning of constructs that exist in the lived experience of people (Havenga, 2008: 13). This is very applicable to this study, as a ‘social' phenomenon, a VRE, is explored in its natural setting.

The approach followed in this study has been qualitative in nature. Qualitative research is described by Creswell (2007: 37) as “the possible use of a theoretical lens, and the study of research problems inquiring into the meaning individuals or groups ascribe to a social or human problem.” The studying of the problem is done through “an emerging qualitative approach to enquiry”, which includes “the collection of data in natural settings sensitive to people and places under study, and data analysis that is inductive and establishes patterns of themes” (Creswell, 2007: 37). This links up to Babbie and Mouton's (2001: 270) view that qualitative research studies human action from the perspectives of the social actors themselves, enabling description and understanding of human behaviour, as opposed to just explaining it (Babbie and Mouton, 2001: 270). Creswell (2007: 53-81) proposes five qualitative approaches to inquiry that can be followed: ethnographic research, grounded theory research, phenomenological research, narrative research, and case study research, which is the approach that this study followed.

The qualitative process can be illustrated as follows (Figure 6.1):

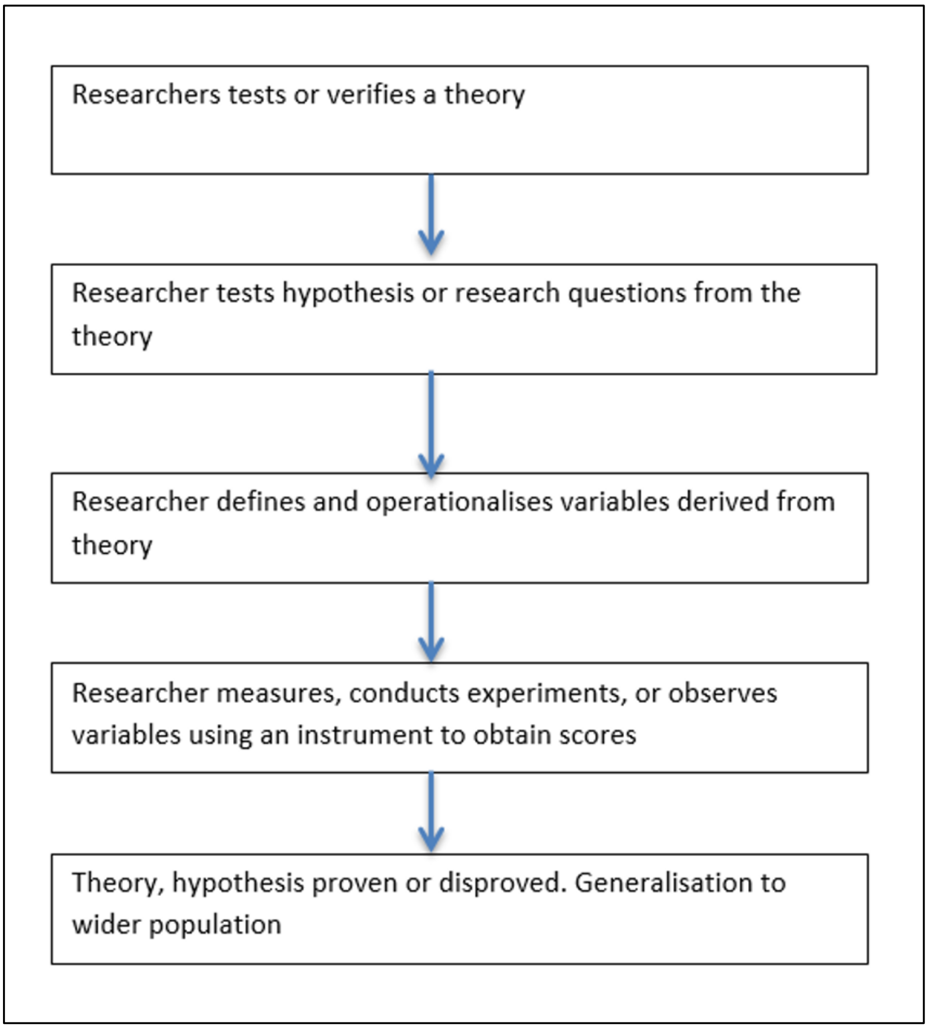
Figure 6.1: The Inductive Logic (Qualitative Research) (Adapted from Creswell, 2003: 132)


Generalisations, or theories to past experience and literature
Researcher searches for broad patterns, generalisations, or theories from themes, or categories generalisation or theories to past experience and literature
Researcher analyses data to form themes or categories
Researcher ask open-ended questions of participants or record field notes (e.g. notes of meetings, notes of observations. e-mails)
Researcher collects information using interview and (participant) observation

Quantitative research, on the other hand, is construed by Bryman (2001: 20) as a “research strategy” focusing on “quantification in the collection and analysis of data.” This is in line with Berg’s (2001: 3) description of quantitative research as the “counts and measures of things” and Creswell’s (1994: 2) description of it as an “inquiry into a social or human problem, based on testing a theory composed of variables, measured with numbers and analysed with statistical procedures in order to determine whether the predictive generalizations of the theory hold true.” According to Bryman (2001: 20), quantitative research follows a deductive approach to the relationship between research and theory, with the emphasis on the “testing of theories.” Quantitative research, as stated by Bryman (2001: 20), has also subsumed the practices and norms that are characteristic of the natural science model, and specifically, positivism. Quantitative research, according to him, also follows objectivism as an ontological orientation. In other words, it views social reality as an external, objective reality (Bryman, 2001: 20). Boutellier et al. (2011: 3) describe the natural science approach to research as being more uniform in nature, relying on more mathematically based methods, countability, and relying to "a large extend on controlled experimental settings." Creswell (1994: 10-11), in turn, identifies two types of quantitative methods, namely experiments and surveys. Boutellier et al. (2011: 3) also mention experiments. Neuman (2000: 121-155) adds content analysis; Graziano and Raulin (2000: 139-146), field work; and Boutellier et al. (2011: 3) systematic observation and measurement.

The quantitative process can be illustrated as follows (see Figure 6.2):

Figure 6.2: Deductive Logic (Quantitative Research) (Adapted from Creswell, 2003: 125)


Researchers tests or verifies a theory
Researcher tests hypothesis or research questions from the theory
Researcher defines and operationalises variables derived from theory
Researcher measures, conducts experiments, or observes variables using an instrument to obtain scores
Theory, hypothesis proven or disproved. Generalisation to wider population

The researcher decided to follow a qualitative approach, as the focus of this study was more to determine through case studies, with the help of observation and interviews, what the essential components are of VREs, and the importance as well as place of RDM in VREs. Quantitative research with its focus on counts and measures would not have yielded the necessary data and results.

The research design used in this study includes both empirical and non-empirical research. Mouton (2001: 51-52) describes empirical research as research that focuses on real-life objects, for example physical objects (matter), biological organisms and processes, cultural objects (art and literature), historical events, social organisations (political parties or clubs), institutions (schools, banks or companies), social interventions (programmes or systems), collectives (e.g. countries, nations or cities), and important for this study, technology, human beings (individuals or groups), and human actions. Non-empirical research, according to Mouton (2001: 52), is research focusing on conceptual problems, for example scientific concepts or notions, schools of thought, philosophies or worldviews, scientific theories and models, scientific methods and techniques, scientific data or statistics, and the body of scientific knowledge or literature.

The non-empirical part of this study consequently consisted of a literature study of the concepts, and the empirical part of the study consisted of real-life objects, in this instance, two VRE case studies and their underlying technology, human beings (members of these groups), as well as human actions.

6.2.1 Literature Review

Pickard (2007: 26) defines a literature review as “a critical discussion of all significant, publically available literature that contributes to the understanding of a subject.” She also identifies four stages in the literature review process: information seeking and retrieval of the appropriate sources, evaluation of sources based on a number of criteria, critical analysis of the content of the literature, and research synthesis, which entails “the ability to synthesize the various concepts and evidence” (Pickard, 2007: 26).

This researcher chose a literature review, because it enables one to discover what is already known about a topic, in this case VREs and their variables. The purpose for doing a literature review was to assist in clarifying the research aims/questions; to supply the necessary depth and breadth of subject knowledge; to provide the theoretical framework for the empirical study; and to contribute to the research design (Pickard, 2007: 25). In other words, this researcher wanted to utilise a literature review as a method to determine what had been done in the field of study by reviewing existing scholarship or the available body of knowledge, so as to get a clear picture of how other researchers investigated the research topic (Mouton, 2001: 87). The literature review also ensured that previous studies were not duplicated, and helped in discovering "what the most recent and authoritative theorising about the subject is" (Mouton, 2001: 87).

The results of the literature review, found in Chapters 2-4 of this study, were used to compile a theoretical framework, which enabled this researcher to propose a conceptual framework/model for VREs, in line with Maxwell's (1996: 25, 37) definition of a conceptual framework. Maxwell (1996: 25, 37) sees it in terms of a concept map, that is, "a visual display of your current working theory," or "a picture of what you think is going on with the phenomenon you're studying." Results of the literature study, as well as the proposed conceptual framework, were then verified by observation of the members in the case studies as well as the findings from the interviews with these members (See Chapters 7-8).

6.2.2 Case Studies

Various attempts have been made in literature to define the concept of a 'case study'. According to Gerring (2007: 19), the word 'case' refers to "a spatially delimited phenomenon (a unit) observed at a single point in time or over a period of time, and includes the kind of phenomenon that an inference tries to clarify. Gerring (2007: 19) further notes that "a case may be created out of any phenomenon so long as it has identifiable boundaries and comprises the primary object of an inference." Rule and Vaughn (2011: 3) corroborate the idea of a case being a spatially delimited phenomenon or unit, when they describe case as "a particular instance." In addition, they add further definitions to the word 'case'. Case, according to them, could also be described as "a circumstance or problem that requires investigation," or as "a body of evidence that supports a conclusion or judgement" (Rule and Vaughn, 2011: 3). The word 'study' is described by Rule and Vaughn (2011: 4) as "an investigation into or of something," which means applying your mind so that you may obtain knowledge. They stress that to study a phenomenon implies a detailed examination of it, often from a variety of viewpoints, in order to get a thorough understanding of it (Rule and Vaughn, 2011: 4). 'Study', according to them, entails some sort of "systematic method that necessitates a depth of examination" (Rule and Vaughn, 2011: 4). A case study is therefore seen by them as "a systematic and in-depth investigation of a particular instance in order to generate knowledge" (Rule and Vaugn, 2011: 4).

Merriam's (2009: 40) contribution to the definition of what a case study is, is her focus on the idea that a case study is a 'bounded system' - a concept she borrowed from Smith (1978: 342). According to her, the "most defining characteristic of case study research lies in delimiting the object of the study," namely the case. A bounded system, according to her, consists of a single entity or unit, which are surrounded by boundaries. In other words, the unit of analysis characterises a case study. Examples of a single case could be a person, a programme, an institution, a community, or a particular policy (Merriam, 2009: 40-41). The fact that unit of analysis (a bounded system) defines the case, makes it possible to combine any or all methods for data collection or data analysis with a case study. Yin (2009: 18), on the other hand, highlights the idea of 'inquiry' and emphasizes that the "boundary between the phenomenon and its context may be unclear" (Runeson et al., 2012: 12). Yin (2009: 18) views a case study as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon and its context are not clearly evident".

Swanborn (2010: 13), emphasizes the notion of a case study focusing on a 'social phenomenon' and its development over a specific period. He describes a case study as "the study of a social phenomenon, carried out within the boundaries of one social system (the case), or within the boundaries of more than one social system (e.g. the cases), for example people, organisations, groups, individuals, local communities or nation states" (Swanborn, 2010: 13). The case study, according to him, takes cognisance of "the case's natural context by monitoring the phenomenon during a certain period, or alternatively, by collecting information afterwards with respect to the development of the phenomenon during a certain period" (Swanborn, 2010: 13). This entails process-tracing, that is, “the description and explanation of social processes” that transpire between participants in the process, or “processes within and between social institutions” (Swanborn, 2010: 13). Initially, the researcher explores that data at the hand of a broad research question, and only after a while develops “more precise research questions” (Swanborn, 2010: 13). Swanborn (2010: 13) suggests a number of data sources, namely “available documents, interviews with participants, and (participatory) observations.” Robson (2011: 136), in turn, in his definition labels a case study a “research strategy” and emphasizes the use of “multiple sources of evidence” (Runeson et al., 2012: 12). Robson (2011: 136) describes a case study as “a strategy for doing research that involves an empirical investigation of a particular contemporary phenomenon within its context using multiple sources of evidence.”

Runeson et al. (2012: 12) propose a definition of a case study for software engineering (a field that relates closely to the concept of VREs), which they derived from Yin’s (2009: 18) and Robson’s (2011: 136) definitions. They describe a case study as “an empirical enquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary software engineering phenomenon within its real-life context, especially when the boundary between phenomenon and context cannot be clearly specified" (Runeson et al., 2012: 12).

Woodside (2010: 1) highlights the idea of deep thinking, deep understanding, and sense making that a case study provides. He proposes a very broad definition of case study research, namely “an enquiry that focuses on describing, understanding, predicting, and/or controlling the individual (i.e. a process, animal, person, household, organization, group, industry, culture, or nationality).” Deep thinking in case study research, according to Woodside (2010: 6), includes knowledge of sense making processes created by individuals (in this case a person, group, or organisation), with sense making referring to how sense is made of stimuli. Sense-making, according to him, enables the individual (person, group or organisation) to focus on and frame what they perceive, and to interpret what they have done (Woodside, 2010: 6). Deep thinking also enables meta-sense making (systems-thinking, policy-mapping and systems-dynamics modelling) (Woodside, 2010: 6). To achieve deep thinking in case study research, Woodside (2010: 6) suggests the use of multiple research methods that will enable triangulation. Methods proposed include direct observation, probing by asking members involved in the case for explanations and interpretations, and analysis of written documents in the case environment (examples of these could be notes taken during meetings, or e-mail communications).

The reason this researcher selected the case study method is the possibility it provides to test the research problems in real-life situations, in this case VREs and their variables, and the possibility of formalising this in a model.

Various types of case studies can be found in literature, for example:

- **Exploratory case studies**
  Exploratory case studies as identified by Runeson and Höst (2009: 135) are in line with Pickard's (2007: 86) "intrinsic case study," carried out for the sole purpose of obtaining a better understanding of the case.

- **Descriptive case studies**
  Case studies for descriptive purposes as identified by Runeson and Höst (2009: 135) are called "instrumental case studies" by Pickard (2007: 86). In these types of case studies, the intention is to explore a specific phenomenon or theory, and the case itself is of secondary importance and becomes more a type of instrument (Flyvbjerg, 2006: 240; Runeson and Höst, 2009: 135).

- **Explanatory case studies**
  Case studies according to Runeson and Höst (2009: 135) can also be used for 'explanatory purposes', for example in interrupted time series design (pre- and post-event studies).

- **Improvement approach case studies**
  Case studies can also be of the type that takes an improvement approach, similar to action research (Runeson and Höst, 2009: 135). This was the approach that was taken in this study.

- **Collective case study**

  The collective case study is a case study where more than one case is used to examine specific phenomena (Pickard, 2007: 86). The latter is the type of case study that this study followed.

Klein and Myers (1999: 69) identify three types of case studies reflecting the research perspective:

- **A positivist case study**

  In a positivist case study, evidence is sought for formal propositions, variables are measured, hypotheses are tested and inferences are drawn from a sample to an identified population. The positivist case study is close to the natural science research model as discussed by Lee (1989: 35-36) and similar to the explanatory type mentioned above.

- **Critical case study**

  The purpose of a critical case study is "social critique and being emancipatory." In a critical case study, different kinds of social, cultural and political domination that may inhibit human potential, are identified. This links up with the improvement approach to case studies mentioned above, because improving case studies may have a character of being critical.

- **Interpretive case study**

  The interpretive case study strives to perceive phenomena through the participants' interpretation of their context, which is in line with the exploratory and descriptive types mentioned above.

The case studies chosen for this study lean towards the positivist perspective, by drawing inferences from the sample to other VREs. The case studies also contain elements of the critical case study and improvement approach case studies, where the aim is to improve the VREs. They furthermore have elements of the interpretive case studies. Two case studies were identified to examine a specific phenomenon, namely RDM at the hand of a VRE (typical of a collective case study).

Pickard (2007: 87) identifies three iterative phases in the case study research process, while Runeson and Höst (2009: 137-138) separate these phases into five separate iterative phases.

The first phase suggested by Pickard (2007: 87) is:

Phase 1 – Orientation and overview

During this phase, the researcher starts with the research question, so as to establish a research focus. Runeson and Höst (2009: 137-138) divide this phase into two. First, they suggest a 'case study design phase,' where objectives are defined and the case study is planned. This is in line with Pickard's (2007: 87) idea of creating a broad aim, with a number of flexible objectives, which are flexible enough to realise the aim(s), but also flexible enough to make room for emerging issues. This helps in setting up the boundaries of the case (Pickard, 2007: 87).

The next step that the researcher must take, according to Pickard (2007: 88), is to decide on whether a single or multiple case design will be adhered to, followed by a selection of a site for the research, which will offer rich and detailed insights. That is, a site that allows for multiple data collection techniques and access to artefacts and people possessing relevant information about the case (Pickard, 2007: 88). Included in this is the signing off and obtaining of permission, in order to establish trust and to build up "rapport with all stakeholders" (Pickard, 2007: 88). This is in line with Runeson and Höst's (2009: 137-138) phase 2, which they identified as the 'preparation for data collection phase'. In this phase, according to them, procedures and protocols for data collection are defined. These data collection procedures and protocols will typically include decisions on the unit of analysis, the type of sampling – which, according to Pickard (2007: 88) is usually purposive sampling in qualitative case study research, setting up a case database to help structure the vast amounts of data gathered, as well as the likely data collection methods that will be followed, for example observations and interviews, which were used in this study (Pickard 2007: 89).

The second phase suggested by Pickard (2007: 90) is:

Phase 2 – Focused exploration

The first step in this phase is identified by Pickard (2007: 90) as data collection. This is where the researcher engages with his/her sample, and starts "gathering data on the case study" (Pickard, 2007: 90). Runeson and Höst's (2009: 137-138) phase 3 is in line with this. They identify phase 3 as 'collecting evidence', which entails the execution of the "data collection on the studied case" (Runeson and Höst, 2009: 137-138). Pickard (2007: 90) stresses the importance of making sure that the data collection technique is appropriate to the research question, and feasible in the context.

Pickard's (2007: 90) second step under focused exploration is called 'iterative analysis'. Case study research, according to her, enables the researcher to refute or confirm emerging themes before vacating the site and can adjust the data collection to respond to emerging themes. In other words, when analysing data, it is important "to be open to all eventualities and not allow prior theory to drive the analysis" (Pickard, 2007: 90). Runeson and Höst (2009: 137-138) name this step phase 4, and call it the 'analysis of collected data'.

The third phase identified by Pickard (2007: 91) is:

Phase 3 – Member checking

In this phase, interviewees are given the opportunity to confirm the credibility of their stories and scrutinise the cross-case themes/topics that have been interpreted by the researcher (Pickard, 2007: 91). Runeson and Höst (2009: 137-138) call this phase the 'reporting phase', or put differently, the feedback phase, which is a better description of what this phase entails. "This feedback is very valuable and sometimes helps see or emphasize something we missed" (Maykut and Morehouse, 1994: 147). A further aspect that can be planned early in the research process is an exit strategy. The reason for leaving the field is normally information redundancy (when collection tools uncover no new information) (Maykut and Morehouse, 1994: 91). Pickard (2007: 92) suggests that the researcher, as part of the exit strategy, give each interviewee at the start of the fieldwork a study outline with dates, even if they are general estimates, so that they can have a clear picture of the outline of the study. After the individual cases have been completed (which includes member checking), the researcher can in the case of multiple case studies move onto cross-case analysis. Finally, the researcher will write up the case study.

For this study, the researcher decided to focus on two case studies (multiple/collective case study method), named Case Study A and Case Study B. These two cases were specifically chosen as each uses different methods in conducting research. Both of these cases started in 2013 as VRE pilot studies at a South African university. Case Study A (consisting of five postgraduate researchers, a promotor acting as VRE manager, and a laboratory manager acting as VRE champion and also co-managing the VRE), uses natural science oriented data and laboratory/experimental methods, whereas Case Study B (consisting of four postgraduate students, a promotor acting as VRE manager, and a librarian) uses human orientated data and survey instruments as data collection method. The researcher of this study as well as a VRE designer were involved in the design of these VREs from the start.

Comparing the use of VREs in the two case studies from different disciplinary settings, was deemed valuable. Pickard (2007: 88) brands the use of multiple studies a ‘collective case study’, although “each case is treated as a single case.” Conclusions derived from each case study are then utilised as data contributing to the study as a whole. Using multiple case studies ensure greater confidence in your findings, a comprehensive understanding of the phenomenon, and also provide the settings “to test the conditions under which the same findings might be replicated” (Paulovich, 2015; Yin, 2012: 7). Multiple case studies also introduce a variety of perspectives and viewpoints that can assist in diminishing the risk of acquiring a “biased point of view” (Stern and Porr, 2011: 51). These two cases were specifically chosen to predict similar results (direct replications) (Yin, 2012: 7). Multiple case studies, according to Pickard (2007: 88), generate an extra layer of analysis, since the individual case studies must be analysed first before any topics can be investigated across and between case studies.

6.2.2.1 Sampling Method

For the purpose of this study, purposive sampling, a qualitative method, was chosen. Purposive sampling, according to Babbie and Mouton (2001: 166), is the method whereby a researcher selects a sample on the basis of his/her own knowledge of the population, its elements, and the nature of his/her research aims. The type of purposive sampling chosen for this study is a priori criteria sampling. In this type of sampling, criteria are identified from the conceptual framework of the research study, providing a broad profile of the characteristics of the participants needed, which can give insight into the major issues of the research (Pickard, 2007: 64). In this study, the researcher identified two case studies working with VREs. Individuals from each of these case studies were identified through purposive sampling and then interviewed and observed. The purposive sampling covered the whole population of these case studies. These included a VRE designer, VRE managers, a librarian, and postgraduate student researchers. The researcher students in each case study were grouped together because they were at the same level of research (all were postgraduate students, working in the same field/discipline, and contributing to the same specific research field, in each of the case studies). The respondents could be divided as shown in Table 6.1.

Table 6.1: Respondents From Case Studies

Case Study A	
VRE designer	1 respondent
VRE manager	1 respondent
VRE champion/co-manager	1 respondent
Researcher students	5 respondents
Case Study B	
VRE designer	Same as in Case Study A
VRE manager	1 respondent
Librarian	1 respondent
Researcher students	4 respondents

6.2.2.2 Triangulation

Yin (2012: 13) describes triangulation as "establishing converging lines of evidence which will make your findings as robust as possible. "The most desired convergence occurs when three (or more) independent sources all point to the same set of events, facts, or interpretations" (Yin, 2012: 13). The purpose of triangulation is thus to gather information from multiple sources and then target it at validating the same facts or phenomenon (Yin, 2012: 92). Within case studies, triangulation can be achieved by using several data collection techniques or numerous sources of evidence, or frequently, both (Pickard, 2007: 86). In this study, observation, semi-structured interviews as well as testing/experimenting were used to collect data. The data collected through these techniques were then triangulated to validate findings.

6.2.2.3 Participatory Action Research (PAR) Method

PAR and case study research can also be considered independent research strategies, but in this study, PAR was used as a data collection method within case study research. The PAR method was chosen to investigate each case study, because it provides for the active participation of the researcher conducting the study, as well as members of the community (case study) under study, "throughout the research process from the initial design to the final presentation of results and discussion of their action implications" (Whyte, Greenwood and Lazes, 1991: 20). In other words, members of the case study under investigation are not treated as passive subjects, but are actively engaged, together with the researcher, in the quest for information and ideas to guide their future actions (Whyte, Greenwood and Lazes, 1991: 20). This correlates well with the case studies explored in this study, where this researcher as well as members of the case studies were involved together in the design of the VREs from their inception to the full implementation.

PAR can be positioned within a social constructivist paradigm (Lincoln, 2001: 130-131; Wimpenny, 2010: 90), as both deal with socially constructed meaning amongst participants. PAR, according to Gergen (1999: 100) and 'Social construction: a reader' (2003: 63) could also be placed within social constructionism, which views the participants that are engaging and making sense of their world from a social perspective, ritual, culture and history (Crotty, 1998: 52-57; Wimpenny, 2010: 90). PAR also links-up with Habermas's (1997: 360-361) idea of communicative action, where participants "find a communicative space where they may find solidarity as understandings of their situation are jointly considered" (Wimpenny, 2010: 90). The origin of PAR stems from social transformation in the developing world, as well as from human rights activism (Fals-Borda, 2001: 29, Wimpenny, 2010: 91). Currently, PAR processes are used to improve situations in business, health, education, social care and community environments (Wimpenny, 2010, 91).

Action research, and also PAR, follows an interpretivist viewpoint of research enquiry, anchored in post-positivist study (Baskerville, 1999: 3-4). The interpretivist viewpoint allows for social intervention in the research situation. The premise is that complex social processes can be studied best by introducing changes into these processes and observing the effects of these changes (Baskerville, 1999: 4). Action research also follows an idiographic viewpoint of research enquiry. In this type of enquiry, the subjects of the study are incorporated as important collaborators into the research. Action research will typically include a team of researchers and subjects as co-participants in the process (Baskerville, 1999: 5).

Due to the multiplicity of fields in which PAR has developed, a variety of definitions exist in the literature to describe PAR (McTaggart, 1991: 169); however, the following definition has been found to be very applicable to this study: PAR is "a philosophical approach to research that recognizes the need for persons being studied to participate in the design and conduct of all phases (e.g. design, execution, and dissemination) of any research that affects them" (Vollman, Anderson and McFarlane, 2004: 129). In PAR, according to Karlsen (1991: 147), a rational democratic dialogue is used as fundamental problem-solving mechanism, and the "most rational solution will probably be achieved in open discussions" where all participants have equal rights. This, however, does not imply "that the researcher should, or can relinquish his or her specific professional contribution and responsibility" (Karlsen, 1991: 148). Some types of questions involve the researcher more than the other participants, such as those related to the "theorizing and knowledge accumulation process itself" (Karlsen, 1991: 149). In other words, the researcher is responsible for reflecting on, and understanding the intricacies of the action process by documenting and analysing it, and to substantiate assumptions about what has taken place through the use of technical knowledge (Karlsen, 1991: 149).

PAR develops through a self-reflective spiral (See Figure 6.3), which can include a spiral of cycles consisting of planning a change with the ‘community’ (group of participants), acting (implementing plans), observing (systematically) the process and outcomes of change, reflecting (evaluating) on these processes and outcomes, followed by more cycles of planning, acting, observing and reflecting (McTaggart, 1991: 175; Wimpenny, 2010: 92). These cycles can be illustrated as cycles that follows each other on a timeline (see Figure 6.3), but in reality, it looks more like a spiral with concentric flows (see Figure 6.4).

Figure 6.3: Participatory Action Research Cycles

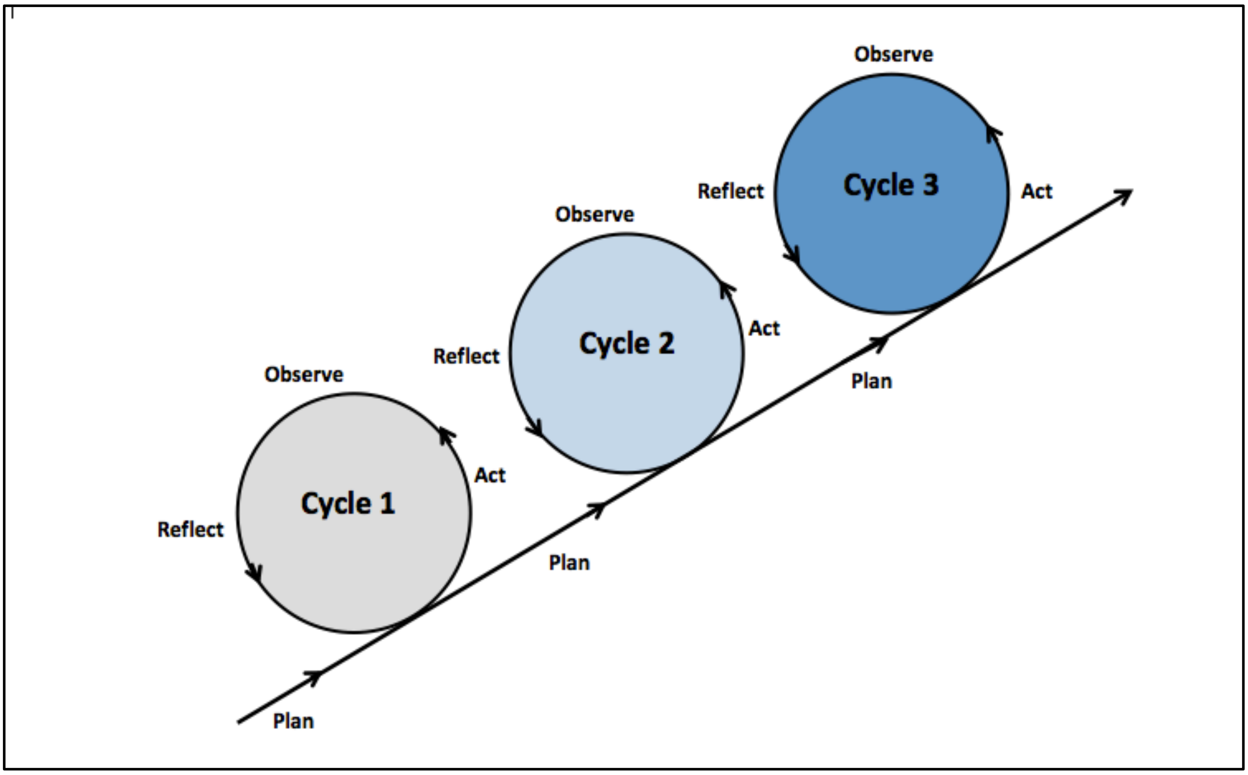
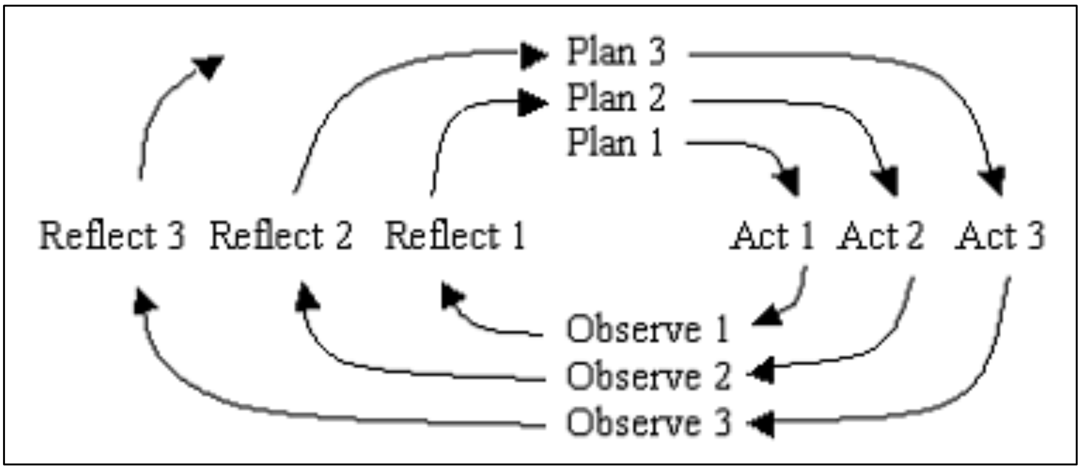


Figure 6.4: Participatory Action Research Model (called an extended action research model by Nguyen, Wegener and Russell, 2006)



6.2.2.4 Prototyping

In conjunction with PAR, prototyping was chosen as a method to develop a VRE for each of the case studies. A prototype is depicted by quite a number of authors (Bischofberger and Pomberger, 1992: 15; Guida and Lamperti, 1999: 3; Hughs and Cotterell, 2002: 6, Lantz, n.d.: 1) as an operational (dynamic and working) model of a system (called a replica by Endres and Rombach, 2003: 10-11), that can be designed, build, implemented and tested by developers and users of the system in order to enable the validation, changing, or refuting or determination of requirements or assumptions. In other words, a model is produced in advance, which exhibits all the essential attributes of the final product, which are then applied as a test specimen and guide for further development (Floyd, 1984: 2). Prototyping is further seen as a quick and inexpensive method to put assumptions to the test (Hughs and Cotterell, 2002: 66).

There seems to be different viewpoints, however, on the degree to which one can simulate the operationality of a system, through a model of a system. Endres and Rombach (2003: 10-11) for example regard prototyping as an activity whereby a partial replica of a system that consist of a subset of functions is set up, in order to determine the requirements for a specific system. Bischofberger and Pomberger (1992: 16) mentioned the construction of 'real prototypes' that have all the important features of the planned system and which can then be used as "specifications for the actual product development process."

The reason for choosing prototyping as a method is because it requires intensive involvement by users (participants) of the system, correlating with the PAR method, resulting in a better clarity of the users' needs and requirements (Moscove, 2001: 68). As mentioned above, prototyping can also be done in a very short time span, which means errors can be detected and eradicated early in the developmental process (Moscove, 2001: 68).

In literature, various types of prototypes can be found. Floyd (1984: 6) identifies three types of prototypes, namely explorative prototypes, experimental prototypes and evolutionary prototypes, which was supported by Bischofberger and Pomberger (1992: 16), while Blomkvist (2014: 28) added evaluative prototyping.

Approaches to prototyping:

- **Explorative Prototyping**

In this type of prototyping, the aim is to clarify requirements and desirable components of the target system (in other words do a requirements analysis), acquire a requirements definition, and to debate alternative probabilities (Floyd, 1984: 6; Bischofberger and Pomberger, 1992: 16). Explorative prototyping is also very useful when the designers of the system "are not domain experts" (Plösch, 2004: 133). This type of prototyping, according to Blomkvist and Holmlid (2011: 4), are predominantly used in the early stages of research, and might "consist of hunches or intuitions that the designer wants to try out." The process typically would commence with initial conceptions of the proposed system, after which a prototype is developed that enables the testing of these conceptions at the hand of real and tangible examples, followed by a consecutive (re)definement of the desired functionality (Bischofberger and Pomberger, 1992: 16). It is also popular in rapid prototyping projects (Blomkvist and Holmlid, 2011: 4). Blomkvist and Holmlid (2011: 4) further emphasise that in cases where the aim is to explore certain aspects or ideas about concepts, the prototyping should be modified "to generate feedback, inspire, and reveal new information." In other words, the prototypes are used as a means of communication, so that the designer(s) can gain a good understanding of the needs of the users (Plösch 2004: 133).

- **Experimental Prototyping**

The aim with experimental prototypes is to obtain an incisive specification of the components of the architecture of the system. This is done by experimentally validating the suitability of system component specifications, architecture models, and ideas for solutions (Bischofberger and Pomberger, 1992: 17). In this type of prototyping, a proposed solution is evaluated through experimental use in order to determine its sufficiency, before investing in a large-scale roll-out of the target system (Floyd, 1984: 8). Experimental prototyping typically begins with the early and inceptive abstraction (conceptualisation) of the different components of the system, after which a prototype is developed to enable the simulation of the designed system components and their interactions (Bischofberger and Pomberger, 1992: 17). This executable model (prototype) according to Plösch (2004: 133), implements (simulates) typical application examples." The individual system components are then tested by the users (Plösch, 2004: 133). Areas that could be investigated include things such as the transparency of the human-system interface, the acceptability of the intended system's performance, and the feasibility of such a solution (Floyd, 1984: 8; Plösch, 2004: 133).

- **Evolutionary Prototyping**

The aim with evolutionary prototypes is incremental system development, in other words the accent is on readjusting the system gradually to changing prerequisites (those user requirements that are evident from the beginning) (Floyd, 1984, 8; Bischofberger and Pomberger, 1992: 17-18). Evolutionary prototyping, according to Bai (2014: 1815), operates iteratively through feedback cycles consisting of design, implementation, and evaluation, and then re-design, re-implementation, and re-evaluation, to accomplish continuous and unforeseeable changes. Bai (2014: 1815) further expands Floyd's (1984) view on evolutionary prototyping by adding an organic view, which includes a sustainable and adaptive 'embryo', which is an organic structure of the future system, as well as embedded feedback management that will enable users of the system (in this study, the members of the two case studies) to communicate with each other and the environment. The evolutionary approach takes the 'embryo' of the system through a process that grows it to a mature system. The outcome then forms the core system for later users, as well as for successive iterative processes during which additional user requirements are integrated. In other words, the prototype becomes the eventual product through an incremental evolutionary process (Bischofberger and Pomberger, 1992: 17-18).

- **Evaluative Prototyping**

Blomkvist (2014: 28) identifies an approach, which he calls *evaluative prototyping*. This type of approach to prototyping is used to “understand how people experience the future that prototyping suggests” (Blomkvist, 2014: 28). This can be formal or informal. Formal evaluation is utilised to test nearly explicit hypotheses or assumptions. On the other hand, informal evaluation is described as “more context-specific and less defined” (Blomkvist, 2014: 28)

The approach followed in this study has been exploratory prototyping. The idea behind this approach is to focus on the communication between the developers (in this instance the VRE designer, as well as the researcher of this study), who are not domain experts, and prospective users (in this instance the postgraduate researchers, VRE manager, VRE Champion, and librarian) early on in the development of an appropriate information technology solution. Floyd (1984: 6) suggests a practical demonstration of possible system functions, which can function as a catalyst to extract good ideas, as well as advocate for an innovative creative collaboration between everyone involved. Plösch (2004: 133) confirms this when he states that “the common technique is to build a prototype that demonstrates the main functionality”, which is then “typically applied during requirements gathering and analysis.”

Floyd (1984: 7) further proposes that when demonstrating the prototype, there should be a strategy in place relating to the choice of components that will be demonstrated. She suggests that it would be useful to allow users to perform one [or more] of their work tasks entirely with the assistance of the prototype. This will enable them to assess the usefulness of it for the specific task at hand.

Hughs and Cotterell (2002: 66) makes a distinction between throw-away prototypes (also called concept protocols by Guida, Lamperti and Zanella, 1999: 10) and evolutionary prototypes (which was also mentioned by Bischoffberger and Pomberger, 1992: 17-18; and Guida, Lamperti and Zanella, 1999: 11). With throw-away prototypes, the prototype is used to try out some ideas and is then discarded when the real construction of the operational system starts. The throw-away prototypes encompass Bischofberger and Pomberger's (1992: 17-18) exploratory and experimental prototypes. Hughs and Cotterell's (2002: 66) definition of an evolutionary prototype also corresponds with Bischofberger and Pomberger's (1992: 17-18) description, mentioned earlier, as well as Guida, Lamperti and Zanella's (1999: 10) description of the aim of an evolutionary prototype, that is, "to reflect user feedback, while evolving the developing prototype to a high-quality usable system that meets user-needs."

Both throw-away and evolutionary prototypes were used in the case studies that this study focuses on, which is in line with Guida, Lamperti and Zanella's (1999: 11) viewpoint that both of these prototypes can be accommodated in the same project.

Guida, Lamperti and Zanella (1999: 10-11) identify the goals of concept prototyping as:

- Determining user requirements;
- Comprehending the functional context;
- Extracting, fine-tuning, and validating detailed prerequisites for human-computer interaction, processing features, storage of data, control of the system, and behaviour;
- Appraising the possibility, proficiency, fitting, or appropriateness of specific solutions to stipulated problems;
- Presenting evidence of feasibility or evidence of a concept;
- Exploring design matters and other possibilities to enhance cognizance of results of decisions;
- Comprehend interfaces with other components;
- Encouraging interaction between the different persons on the design team and between designers (developers) and users;
- Sanctioning designers (developers) and users "to learn incrementally as the project evolves", because several requirements, restrictions and aims of the system being designed, are undetermined at the start;

- Presenting a method to fashion a final software system that fulfils the real needs of users, minimise post-delivery costs, and increases quality; and
- Limiting forfeiture if the project is cancelled.

With regard to evolutionary prototyping, Guida, Lamperti and Zanella (1999: 12) add some additional goals with respect to the concept prototyping:

- Procuring reliable user feedback on the real user interface and the outputs of the system that is being constructed;
- Finding solutions to performance issues and assisting in implementing them; and
- Contributing value to the target group in a timespan.

The implementation of an evolutionary prototype normally occurs in an environment that is suitable for a quick build and rapid refinements. This normally happens on the target platform, but if this is not possible, it is done in such a manner "that cross-platform portability is guaranteed" (Guida, Lamperti and Zanella, 1999: 15). When the prototype reaches a certain level of sophistication, it is implemented as a pilot system (Lichter, Schneider-Hufschmidt and Züllighoven, 1994: 826) as the core of the ultimate system. In other words, the prototype then becomes the application system.

According to Moscove (2001: 67), prototyping consists of four major development stages: identify information system requirements, develop initial prototype, iterative process, and use of approved prototype. Each of these steps is described below.

(a) Stage 1 - Identify Information System Requirements

This stage correlates with Floyd's (1984: 4) 'functional selection' step, which pertains to the choice of functions that the prototype should display. This choice, according to Floyd (1984: 4), should always hinge on relevant work tasks, which can function "as model cases for demonstration." This is the stage, according to Moscove (2001: 67), where developers meet with the users of the system to determine the user's expectations or requirements for such a system.

(b) Stage 2 – Develop Initial Prototype Construction

This stage correlates with Floyd's (1984: 4) 'construction' step, which is concerned with the exertion needed to make the prototype available. During this step, an initial prototype that meets the fundamental requirements, as identified in the first stage, is developed (Moscove, 2001: 67). This is then followed by a demonstration of the prototype to the users of the proposed system for the purpose of experimentation. The users then list their opinions and experiences, as well as their recommendations about the system (Moscove, 2001: 68).

(c) Stage 3 – Iterative Development Process

Following the second stage, the system developers make the necessary adjustments and changes to the initial prototype. This is done through a process of formative evaluation, a continual process before and during the implementation of a programme or system (in this study, a VRE technological framework), and summative evaluation, a process that is done after the implementation, to determine the impact it had if its implementation was successful (See 6.5 for more detail on these two types of evaluation) (Floyd, 1984: 4; Scriven, 1991: 168-169; Evaluation Toolbox, 2010).

A document containing the explicit criteria for evaluation should form the basis for the evaluation. This document should also stipulate the steps to be performed within the system. Evaluation may occur at the level of single users using the system, in which instance the focus is normally on cognitive problems that involve the machine-user interface, or it may concern the co-operation between participants, and participants and other people, which will require a scrutinisation of communication between people (Floyd, 1984; 4-5). After this, the revised prototype is again presented to the users to experiment with (Moscove, 2001: 68). This iterative process is repeated, until the users are satisfied with the changes to the system. The iterative nature of prototyping makes it naturally an ideal method to apply in conjunction with action research, which also follows an iterative process.

(d) Stage 4 - Use Of Approved Prototype

During this stage, the approved prototype is turned into a fully functional information system, called an operational prototype by Moscove (2001: 68). Possibilities for further use depend on the experiences acquired with the prototype, and on the available production environment. It may just function as a learning tool and be thrown away thereafter, or it may be implemented fully or partially (Floyd, 1984: 5).

This researcher followed Paulovich's (2015) process for prototyping. First was the development of the initial design concepts, which were reviewed at regular intervals by members of these case studies, to ensure that the concepts were appropriate, accurate, constructive, and useful. Protototyping was also used during these two projects to evaluate the evolving design and to obtain perspectives from members of these groups. This design process followed a cyclical path, which is continuing until the prototypes have been fully implemented as working VRE technological frameworks.

6.2.2.5 Data Collection Protocol

Various techniques for data collection can be used in PAR, depending on the issue or situation (MacDonald, 2012: 41); however, for this study, the researcher chose two types that would be suitable. Two techniques were chosen in order to transcend the limitations of each one, and to triangulate data generation and ensure the best validated results or findings (Winter, 1989: 22; Streubert and Carpenter, 1995: 257, 318). The two techniques are: interviews and participant observation (notes of meetings with researchers, which include formal and informal documentation, as well as informal communication with researchers via e-mail). The data collection methods used for prototyping consisted of testing and experimentation.

(a) Participant Observation

Participant observation affords a researcher the opportunity to have access to research subjects in a social setting, and encapsulates the context of the social environment in which individuals operate by documenting human behaviour (Gillis and Jackson, 2002: 210, 229-230; Mulhall, 2003: 308, 310). In this method, the researcher as participant-observer collects data in a "relatively unstructured manner" by observing participants, activities, and facets of the situation, while also participating in activities, rituals, interactions and events of the people being observed (Dewalt, Dewalt and Wayland, 1998: 260). Practically, this technique requires the systematic noting and recording of behaviours, events, and objects in the social situation (Marshall and Rossman, 2006: 98).

The reason for using this technique was to observe how members of these case studies (VREs) behaved and interacted with the various components in the VREs (Pickard, 2007: 201). In other words, the researcher watched what they did, then recorded this in notes of meetings and e-mails, and described, analysed and interpreted what he had observed (Robson, 1997: 190). A semi-participant observation technique was followed in this study. In this type of technique, this researcher watched, interacted and recorded, but kept interaction to a minimum, typically only asking questions to substantiate what had been seen (Pickard, 2007: 204).

(b) Interviews

Interviews usually take the form of a conversation between the investigator of a study and the individuals identified as part of the sampling. The purpose of an interview is usually to acquire "qualitative, descriptive, [and] in-depth data that is specific to the individual" and the case study (Pickard, 2007: 172). Interviews provide the means to discover individual opinions, and provide interviewees as well as the interviewer "with the opportunity to clarify meanings and share understanding" (Bertrand and Hughes, 2005: 74; Pickard, 2007: 172).

There are various types of interviews that can be considered, for example, basic individual interviews, in-depth individual interviews, very structured formal interviews, semi-structured interviews, and focus group interviews, etc. (Babbie and Mouton, 2001: 289-293). This researcher chose semi-structured interviews as an instrument, because its face-to-face interaction between the interviewer and an interviewee offered an understanding of experiences or circumstances as described by the interviewee in his or her own words (Schurink, 1998: 20). Semi-structured interviews can also be called guided interviews (Tutty et al., 1996: 65). The flexibility inherent in these types of interviews made it possible for the researcher "to explain questions and to elaborate on them" (Van Wyk, 2005: 5). It further allowed the researcher to explore unplanned topics arising during the interviews, and enabled him to understand the interviewees' viewpoints and reasons behind them (Van Wyk, 2005: 5). Bryman (2001: 315) also suggest semi-structured interviews when dealing with multiple-case study research, because of the structure they provide for cross-case comparability, as in this case, where the study focussed on two case studies.

Recording medium for the interview

The researcher decided upon audio-recording of the individual interviews for internal analysis only. Permission was obtained from the interviewees (respondents) beforehand and in cases where the respondents were reluctant to be audio-recorded, written notes were taken of the interviews. In the case where a respondent was not geographically present, the researcher made use of Skype as a medium to conduct the interview and record the interview. Each audio-recording was time- and date stamped and the interview duration was also indicated. The audio-recordings were then transcribed by extracting only the main points from the conversations. Thereafter, the individual transcriptions were sent to each of the respondents to verify if the transcription of their answers were correct.

Interview Schedule

The semi-structured interview usually consists of a list of questions or relatively specific topics to be covered, which are prepared beforehand by the researcher. These are normally described as an interview guide or interview schedule. As mentioned, the semi-structured interview is much more flexible, which gives the interviewee much more freedom in how to reply, which means questions may not exactly keep to the path defined in the schedule. During the interview, the researcher might pick up further information, which might lead to further questions not included in the original schedule, but he/she will mostly guide the interview in such a way as to keep as close as possible to the original schedule, and make sure that all questions outlined in the schedule are asked, and that similar wording is used with each respondent (Bryman, 2001: 314).

Bryman (2001: 317) suggests some basic elements in the preparation of an interview guide/schedule:

- Construct some order in the topic areas, so that the questions on them flow quite well, but be ready to change the sequence of questions during the actual interview;
- Frame interview questions or topics in a manner that will enable the solving of research questions/problems;
- Use language that is understandable and relevant to the interviewees;
- Do not ask leading questions;
- Document 'face sheet' information, e.g. general info (name, age, gender, etc.) and specific info (position in company/organisation, total years unemployed, total years involved in a group, etc.) – this information would be valuable for contextualising the responses. However, in this study, the age, gender, total years of employment, and position in the organisation were not included in the face sheet, as they were not relevant to this study. It also had ethical implications in that it could make it possible to trace the results of the interviews back to certain individuals.

The above basic elements of an interview schedule as suggested by Bryman (2001: 317) were taken into consideration when an interview schedule for this study was compiled. The questions (see 6.3) were framed in such a manner that they would address the research sub-questions/problems. They are ordered in four sections. The first section contains questions that touch on the background/context of the case study; the second section has questions dealing with the VRE itself; the third section has questions relating to RDM; and the fourth section includes general questions on the VRE. During the interview, however, the researcher allowed for flexibility to ask further questions, when necessary, in order to prompt the respondent(s) for more information. This meant, in some cases, that the questions did not exactly keep to the path defined in the original schedule.

As part of each interview, the researcher also completed 'face sheet' information for each interview. This included the following:

- Date of interview;
- Place of interview;

- Name of respondent/interviewee (although this was asked, it was only for follow-up purposes; the name was not used in the study, for ethical reasons);
- Name of interviewer;
- Name/Title of research (if applicable);
- Position/role in the project;
- VRE project name;
- Total of years/months involved; and
- Duration of interview.

The sequence of each interview can be summarised as follows:

- Introduction: As an introduction to the interview, the researcher thanked each respondent for the possibility to interview, and explained the reason for the interview. This was followed by reading, discussing, and signing the confidentially agreement with the respondent. Finally, a timeframe for the interview was agreed upon;
- Completion of face sheet information; and
- Questions listed in 6.3.

6.2.2.6 Testing And Prototyping

The tools used as technological framework (user interface) in the case studies, were iteratively designed through a process of testing and prototyping. "Iterative design of user-interfaces involves steady design refinement based on user-testing" and prototyping (Nielsen,1993: 32; Interaction Design Foundation, 2017). The iterative process is frequently called "rapid prototyping or spiral prototyping" and usually ends once you have reached the best possible technological framework for use by the target group (Interaction Design Foundation, 2017). The benefits of using an iteration design approach are described by Interaction Design Foundation (2017) as follows:

- It enables "rapid resolution of misunderstandings" and brings clarity;
- Allows for user feedback, and ensures that user needs are met;
- Improves client relationships through showcasing the "evolution of a design";
- Ensures that the design team's efforts are focused on adding value for the users;
- Makes it possible to incorporate lessons learned in the final product;

- Provides opportunities for regular testing, which can then yield a powerful performance framework that can be used for “acceptance testing”; and
- Increases visibility of the progress at every iteration.

The iteration process could typically go through the following sequence:

Step 1: Plan the design of an initial interface through user requirements gathering;

Step 2: Complete an initial interface design;

Step 3: Present the design to several test users;

Step 4: The users test the design;

Step 5: Evaluate the problems the test users encounter when using it;

Step 6: Improve the interface by analysing flaws and problems, and redesign the interface; and

Step 7: Repeat the iterative process until all the user-interface problems / flaws are resolved (Nielsen, 1993: 32; Interaction Design Foundation, 2017).

This iterative design process has been applied by the researcher of this study in the design of a VRE interface for two project groups (case studies) through a process of formative evaluation. The iterative process followed included the following (see also 7.2.1 and 7.2.2):

Step 1: Identifying a case study;

Step 2: Exploring the case study;

Step 3: Expanding the case study through a needs identification (requirements gathering) and developing a prototype;

Step 4: Demonstrating the initial prototype to the users (presentation of the design to several test users);

Step 5: Users tested the prototype and gave commentary on the prototype;

Step 6: Evaluation of users’ problems and adapting (redesigning) the VRE system;

Step 7: Users tested the system again, they provided feedback (commentary), and more needs were identified and evaluated. This led to a further adaption (redesign) of the VRE system. The redesigned system was then implemented. This iterative process was repeated until the users were satisfied with the system's functionalities.

6.2.2.7 Data validation

Data validation according to Informatica (2018) is "a means of checking the accuracy and quality of sources data before using, importing or otherwise processing data." In this study, after the interviews were done, the researcher sent transcriptions of the individual interviews to the individual interviewees for clarification, validation and commentary. The data was then corrected and adjusted in line with recommendations from these respondents.

6.2.2.8 Ethical considerations, and data confidentiality, access and use

The nature and objectives of the study was explained to each of the respondents, and an informed consent form was signed by each, which gave the researcher permission to use the results from this study for the purposes of publication. The researcher of this study also signed a declaration that information / data obtained through the interviews would be handled confidentially. Information of respondents were anonymised through the use of coding. The raw anonymised data will be kept in an institutional data repository or archive after publication. It will be made available for re-use and sharing.

6.3 AN OVERVIEW OF THE QUESTIONS DEALT WITH DURING THE INTERVIEW

In Chapter 1, the researcher listed the research problem/question as well as its sub-questions. These were:

Research Problem / Question:
How can a Virtual Research Environment be conceptualised to indicate the role of Research Data Management (RDM) within a VRE?

Sub-questions:

- What is a VRE?
- What is the current state of VRE research in the world?
- What are the generic components that make up a VRE?
- How does a VRE support a research cycle?
- What is RDM?
- Why should a VRE be an essential technological and collaborative framework for the management of research data?
- To what extent can the components identified in the third sub-question be formalised into a conceptual framework?
- Where would RDM as component be placed?
- To what extent can this model be generalised for use in other environments?
- How was the central research question answered?

These questions were taken into consideration in the drawing up of questions that would be asked during the interviews with members of the VRE case studies. The questions were divided into two units: the face sheet with background information (i) and questions asked (ii). The questions were divided into four parts. Part 1 includes questions to postgraduate researchers participating in the VRE. These questions were further divided into three sections: Section A deals with questions on the VRE, Section B deals with questions on RDM, and Section C deals with general questions on the VRE. In Part 2, the researcher included specific questions for the VRE managers; Part 3 included specific questions to the VRE designer; and Part 4 included specific questions for the information specialist/librarian.

6.3.1 Face Sheet (Background Information)

The information in the face sheet was collected by the researcher for administrative purposes (e.g. follow-up with respondents) only and were not used as part of the empirical process or analysis of results. The face sheet asked the following:

- Date of interview;
- Place of interview;
- Name of respondent/interviewee;
- Name of interviewer;

- Position/role in the project;
- VRE project name (Project 1 or Project 2); and
- Duration of Interview.

6.3.2 Questions

6.3.2.1 Part 1: Questions To Student Researchers Participating In The VRE

Section A: Questions on the Virtual Research Environment (VRE)

(1) This VRE project exists 18 months, when did you join this project? (Number of months)

The members of the two VRE projects typically consisted of researchers that were at different stages in their postgraduate degree studies, which means that some of the members in these projects would be more familiar with the tools of the VRE system, than others. This question aimed to determine the length of time a respondent has been involved in the VRE. This could potentially influence the answers to the rest of the questions.

(2) What is your role within the VRE project?

Different roles were assigned to each of the members of the projects. This question was asked to determine whether the respondent was a student researcher, information specialist (librarian), an IT specialist (VRE designer), a lab manager (VRE facilitator), or a VRE project manager.

(3) Does the VRE project you belong to, focus on one discipline, or would you say it is multidisciplinary?

This question was asked to determine if the members of a VRE were part of one research field or were from multiple disciplinary areas. The needs of the different disciplines might differ with regards to a VRE.

(4) *Is the focus of the project topic-centred or technology-driven? Please explain?*

This question was asked to determine if the project was technology driven or driven by needs of the research project itself.

(5) *What do you understand under the term "Virtual Research Environment" (VRE)?*

This question was asked to determine if the respondent realised what a VRE is, and if his/her ideas corresponded with what the literature review showed.

(6) *Were you afforded an opportunity to give input in the design of the VRE? If so, what type of input did you give? (If you joined the VRE project after it was launched, were you afforded an opportunity to comment on the current design?)*

This question was asked to establish if the VRE projects were implemented in a top-down approach or if the respondents could give input on how these VREs should look like.

(7) *Did you receive training to use this VRE? If so, what type of training did you receive?*

The researcher wanted to determine through this question, the user-friendliness of the VRE and the know-how of the researchers using the VRE. A respondent that did not attend the training might have experienced the VRE differently from respondents that did.

(8) *Which of the current tools/components in the VRE are you currently using? Please indicate why you are you using each of these tools/components, and not the others?*

- Create a Site □ ____________________________________
- Edit your profile □ ____________________________________
- Search □ ____________________________________
- Site Calendar □ ____________________________________
- My Discussions □ ____________________________________

- Following □ ______________________________
- My Files (Drag and Drop, Upload Files, Create a Folder) □

  ____
- My Activities (News) □ ________________________
- Site Activities □ ____________________________
- My Tasks (Task assigned to you – Workflow function) □

  ____
- My Documents (Keeping track of your own content) □

  ____
- Shared Files (Files that everyone has access to) □

  ____
- People Finder □ ____________________________
- Invite Users □ _____________________________
- Discussions □ _____________________________
- Document Library □ _________________________
  - Categories □ _________________________
  - Tags □ ______________________________
  - Favourite □ ___________________________
  - Like □ _______________________________
  - Comments □ ___________________________
  - Share □ ______________________________
  - Edit Properties □ ________________________
  - Edit Offline □ __________________________
  - Dublin Core Metadata Template □

    ____
  - Manage Permissions □ _____________________
  - Upload New Version □ _____________________
  - Download function □ ______________________
- Instrument Backups □ ________________________
- Software Backups □ _________________________
- Survey or Questionnaire Tool □ ____________________
- Publishing Function □ ________________________
- Mobile syncing with Alfresco database □ ________________
- Desktop syncing with Alfresco databas □ ________________

This question was asked to determine the uptake of the various components in the Alfresco system and why some was used and others not

(9) *Do the tools available in the VRE meet all your expectations? If not, why not?*

Through this question, the researcher aimed to discover needs with regard to components in the VRE.

(10) *Do you have someone from your research group/department that act as champion/facilitator for the group? What is his/her designation/title in the group?*

In 3.5.7.1, it was shown that the VRE facilitator plays an important role as one of the human components of a VRE. This question was aimed at determining if the VRE had a facilitator and what his/her role was.

Section B: Questions on Research Data Management

(11) *What would you describe as research data? Why do you see some data as not being research data?*

The purpose with this question was to gauge the respondent's knowledge about research data, and what he/she did not consider as being research data. This links up with the distinction that was made in 4.2.1 between research data, referencing data, funding data and collaboration data.

(12) *How would you describe the concept "Research Data Management"?*

This question was aimed ascertaining what the respondent understood under the concept RDM. His/her answer(s) to this question would have a direct reflection on the answers to further questions.

(13) *To what extent would you say that research data could be managed through a VRE?*

This is a very important question, which touches on the core of the research problem this study aimed to address. It was aimed at eliciting ideas from the respondents on how research data could be managed by means of a VRE. This would then be correlated with ideas from the discussion in 5.3 and the conceptual model in 5.4.

(14) *How did you manage your research data before becoming part of the VRE project? How are you using the VRE to manage your data?*

This question was aimed at determining RDM methods used before the respondent became part of the VRE, and how the VRE added value to the respondent's research workflow.

(15) *Are you able to do the following RDM related tasks/actions within the VRE: (Please explain how it is done and if not, why not)?*

a. Create/capture data, using:
   - Major instruments (e.g. telescopes, accelerators, specific software programmes, sequencing)
   - Simulations
   - Laboratory experiments
   - Surveys
   - Literature
   - Other

b. Store/backup data

c. Store different versions of data

d. Add metadata. If so, what metadata are needed for technical and scientific reasons? Who adds the metadata?

e. Process data

f. Analyse data

g. Visualise data

h. Share data, with peers or with supervisor (Workflow)

i. Publish data in a repository. If not, are you publishing your data elsewhere?
j. Preserve data for long-term

This question was specifically focused on RDM and the various components and processes of the research data lifecycle, and aimed to discover which of these were functioning through the two VREs.

(16) *Is the use of the data restricted by the following:*

a. Confidentiality/ethical reasons;
b. Law;
c. Proprietary / commercial interests;
d. Creative Commons License; or
e. Other?

The aim of this question was to determine the openness of the data that were managed. This would also influence the possibility of publishing the data in a repository, as well as the licensing of the data.

(17) *What methods/tools do you use to analyse the data?*

The researcher purposed with this question to find out which data analysis tools were used by researchers in the research process/cycle.

(18) *Do you use visualisations to analyse your data? If so, give examples.*

Visualisations are used more and more as a method to analyse data. This question tried to determine the use of visualisations, if any.

(19) *Are RDM-related issues formalised within your VRE? (For example, Strategic action plans). If yes, to what extent do you adhere to the policy?*

Strategic action plans with regard to RDM are normally formalised in a DMP or a policy / strategic document. The purpose with this question was to determine if these VRE case studies had such documents in place.

Section C: General Questions on the VRE

(20) *Describe what other functionalities you can use in the VRE. Are you utilising these? If so, which functionalities? If not, why not?*

This question intended to determine the awareness of the respondents about the various functionalities/tools of the VRE. It was also asked to find/gauge the uptake of these functionalities/tools.

(21) *What would you say are the objectives of the VRE within which the research project is managed?*

a. Immediate objectives.
b. Is there a time limit for the VRE?
c. Objectives beyond the project period.

This question and sub questions aimed to find out the reason for the existence of the groups, the life-span, and sustainability, as well as what the long-term plans were after the project was discontinued.

(22) *Has the use of the VRE benefitted your research/work processes? If so, how? If not, please explain.*

This question was asked to establish the value(s) that the VRE had for the respondents.

(23) *Were there any obstacles in using the VRE? Would you suggest any changes?*

The purpose with this question was to determine any shortcomings with regard to the VREs, which could be added to the conceptual model.

6.3.2.2 Part 2: Questions To VRE Managers

Some of the questions in Part 1 were applicable to the VRE managers, but not all. In the interviews with the VRE Managers, the relevant applicable questions from Part 1 were selected, together with the questions in Part 2, which specifically focused on VRE managers.

(24) How and when did the VRE project to which you belong, start and develop?

The researcher included this question in order to ascertain what processes were followed to create and develop the VRE projects.

(25) What are your tasks as VRE Manager?

This question was asked to find out what the VRE Manager's role responsibilities were, and whether a VRE Manager is an essential human component in a VRE.

(26) How do you ensure that the members stay engaged in the VRE?

It is essential for the success of a VRE that people use the VRE. This question was asked to determine the methods the manager used to keep the members engaged.

(27) How do you handle technical problems that surface in the VRE? Give examples of how such problems were addressed.

This question was asked to determine if technical problems in the VRE were fed back to the VRE designer, and if the VRE was adjusted accordingly.

(28) How do you address additional needs in the VRE? Give examples.

This question was asked to discover if there were feedback to the VRE designer about additional needs and if the VRE was re-evaluated and adjusted.

(29) *How do you define the added value of the VRE for the members of the VRE?*

This question touched on the tasks/role of a VRE manager, and was asked to establish what role the manager played in ensuring that members understood the added value of the VRE.

(30) *How do you ensure quality control in the VRE?*

With this question, the researcher wanted to probe what quality control measures were in place.

(31) *Does the project have a formal RDM strategy/plan? If yes, please elaborate. If not, please explain.*

An RDM strategy/plan is important to structure and guide RDM activities in a VRE. This question aimed to establish if such a strategy/plan was in place, and what it entailed.

(32) *If the project does have a formal RDM strategy/plan, how do you ensure compliance within the group?*

This question links up with the tasks/role of a VRE manager; it was asked to determine how the manager as part of his/her role ensured that the strategy/plan was adhered to.

6.3.2.3 Part 3: Questions To VRE Designer

Some of the questions in Part 1 were applicable to the VRE designer, but not all. In the interviews with the VRE designer, the relevant applicable questions from Part 1 were selected, together with the questions in Part 3, which specifically focused on the VRE designer.

(33) *What are your tasks as VRE designer?*

The intention with this question was to establish the role of, and list the responsibilities the designer has in a VRE.

(34) *What is the process you followed to design these two VRE projects? Did you first create a prototype for the VRE(s)?*

The purpose with this question was to establish what the steps in the design process were that were followed to construct the final product.

(35) *What software(s) did you use to design the VRE?*

This question was aimed at determining the various software the designer used in the design of the VRE(s).

(36) *How did you decide upon the specific software(s)? Why did you use this specific software(s)?*

With this question, the researcher aimed at establishing the reasons for using this software(s).

(37) *How did you determine which functionalities (components) the members of the VRE groups needed in the VRE?*

This question was asked to ascertain if the members of the VRE was consulted in the design of the VRE.

(38) *What are the functionalities that you made provision for?*

This researcher wanted to extract a list of potential VRE components with this question.

(39) *What are the hardware and software infrastructure specifications for RDM activities within your VRE? (e.g. storage and computing capacity needed?)*

The question was asked to determine the size and format types of data that the system needed to accommodate. This would give an explanation of the hardware and software used.

(40) *How did you ensure that the data in these VREs are protected from loss or damage?*

This question was aimed at discovering the actions taken to ensure the security of data in the VRE(s).

(41) *Do the VRE(s) systems make provision for data publishing as well as long-term preservation of data?*

This question was asked to determine if the VRE provided for other aspects of the research data cycle such as publishing and long-term preservation of data.

(42) *Did you have to make any adjustments to the VRE? If so, what did you do?*

With this question, the researcher wanted to determine if there were revisions made to the VRE(s), and what these were. This connects to the method of action research and prototyping.

(43) *What type of training, if any, did you give to the students, researchers, librarian, and VRE managers?*

This researcher asked this question to confirm the answers given by other members of the VRE(s) in question 14.

(44) *What future developments do you envisage for the VREs?*

This question was asked to establish what future developments were planned for the VREs.

6.3.2.4 Part 4: Questions To Information Specialist / Librarian

Some of the questions in Part 1 are applicable to the Librarian, but not all. In the interview with the Librarian, the relevant applicable questions from Part 1 were selected, together with the questions in Part 4, which specifically focussed on the Librarian.

(45) *Do you think a librarian has a role to play in a VRE? If so, what do you see as the potential role(s) a librarian can play in a VRE?*

This question was asked to ascertain whether the librarian understood what the potential role of librarians are in a VRE, and what the role is.

(46) *What are your tasks as librarian in this VRE?*

This question links up with the previous question and was asked to obtain a list of tasks and potential tasks a librarian could perform in a VRE.

(47) *Do you have any specific role in terms of RDM in the VRE? If yes, what? If not, why not?*

This question was asked to determine what role, if any, a librarian can play with regards to RDM in a VRE.

(48) *How did you get involved in this VRE(s)?*

The researcher aimed to discover with this question, how librarians can be enticed to get involved in VREs.

(49) What do you see as the value of a VRE for you as a librarian?

With this question, the researcher wanted to determine the value that a VRE can have for librarians.

6.4 METHODS OF ANALYSIS

Methods used to analyse the results gained through the observations and semi-structured interviews, included the following:

- Pattern-matching: patterns emerging from the data collected from the case studies, were matched with patterns found in the results from the literature study. A pattern in case study research is described by Almaturi, Gardner and McCarthy (2014: 239) as “an arrangement of occurrences, incidents”, behavioural actions, “or the outcomes of interventions” that can be found in the raw data. The value of using pattern-matching lies in “its ability to link research data” flowing from the interviews, “with the theoretical proposition”, which can be gained from prior research, knowledge, or theory as found in literature (Almaturi, Gardner and McCarthy, 2014: 242). The aim with pattern-matching is “to build explanations on whether or why the patterns are matched or not”, which eventually leads to greater validity that supports or modifies the theory, or in this case a conceptual model, that underpins the study (Yin, 2003 as cited by Almaturi, Gardner and McCarthy, 2014: 242).
- Analysis of tools, where tools used by participants in the two case studies mentioned, were analysed qualitatively using a conceptual framework of specific criteria that emerged from the literature study (See Figure 3.12a-c).

The above methods of analysis helped in the testing of findings for their fit with previous research and theory on the subject. Linkages between findings and previous knowledge helped to demonstrate the generalisability of the findings, called *analytic generalisation* by Babbie and Mouton (2001: 283).

6.5 EVALUATION

Two types of evaluation are identified in literature, namely formative and summative evaluation.

6.5.1 Formative Evaluation (Interactive, Informal)

Formative evaluation, according to Scriven (1991: 168-169) and Evaluation Toolbox (2010) is typically performed before or during the implementation, “development or improvement of a program or product (or person, and so on) and it is conducted, often more than once,” to improve the programme or product’s design and performance. Formative evaluation complements summative evaluation by striving to comprehend why a programme works or not, as well as what other factors (internally or externally) are operating during a project’s lifespan (Evaluation Toolbox, 2010). This type of evaluation is well-suited for the evaluation of the VRE case studies, as they developed iteratively through improvements. The Evaluation Toolbox (2010) provides a valuable table based on Owen and Rogers (1999: 39-62), which lists the various categories/dimensions and actions of formative evaluation. The table have been adapted into Table 6.2, to indicate which data collection techniques were used in each stage of this study.

Table 6.2: Formative Evaluation (adapted from The Evaluation Toolbox, 2010)

	Pro-Active Dimension	Clarificative Dimension	Interactive Dimension	Monitoring Dimension
When	Pre-project	Project development	Project Implementation	Project Implementation
Why	To understand and clarify the need for the project (Needs assessment).	To make clear the theory of change that the project is based on (clarification of project design).	To improve the project’s design (continual improvement as it is rolled out).	To ensure that the project activities are being delivered efficiently and effectively (fine tuning).
Techniques used in this study	Meeting with decision makers, and potential participants (Documents of meetings and copies of e-mails).	Regular Meetings with VRE users, E-mails (Documents/Notes of meetings and copies of e-mails).	Regular Meetings with VRE users, E-mails (Documents/Notes of meetings and copies of e-mails).	Regular Meetings with VRE users, E-mails (Documents/Notes of meetings and copies of e-mails).

(a) Pro-Active Dimension

Evaluation in this dimension usually takes place before a project or programme is designed (Owen and Rogers, 1999: 41). This dimension or stage is used to do an analysis of the needs of a group, and to determine what type of (if any), programme should be designed/developed to meet these needs (Owen and Rogers, 1999: 41). Techniques used for this study included a needs assessment through meetings on site with decision makers and potential participants, as well as review of documents and e-mails related to the project.

(b) Clarificative Dimension

In this dimension, the internal structure, rationale and functioning of a programme or platform are evaluated (Owen and Rogers, 1999: 42). This typically leads to modification of elements in the programme in order to address the intended outcomes of the project (Owen and Rogers, 1999: 53). Techniques that were used for this study include meetings with participants, analysis of notes taken during these meetings, and e-mails, as well as observation of the usage of the programme or platform.

(c) Interactive Dimension

The interactive dimension renders information about the implementation and execution of a programme or platform, or specific components or elements of it (Owen and Rogers, 1999: 44). This type of evaluation is valuable in programmes that are continually evolving and improving, and is typically used for incremental improvement of the programme as it is rolled out (Owen and Rogers, 1999: 44). Data collection techniques used for this study included regular meetings, notes taken during these meetings, and e-mails, as well as observation of the programme use.

(d) Monitoring Dimension

The monitoring dimension is used when the programme or platform is well-established and continuing, and is employed to determine if the identified programme targets and implementation are taking place (Owen and Rogers, 1999: 46). Issues that are looked at are efficiency and effectiveness of the implementation and ways to fine-tune the programme to make it more efficient and effective (Owen and Rogers, 1999: 53). A broad range of data collection techniques could be used to analyse the performance of a system and its components, but in this study, the following techniques were used: regular meetings, notes taken during interviews, feedback from users of the system, and e-mails.

6.5.2 Summative Evaluation (Formal)

Summative evaluation according to Smith (2012: 173) "provides information on a product's efficacy (its ability to do what it was designed to do)." According to Evaluation Toolbox (2010), summative evaluation examines the impact that an intervention, in this case the design of a VRE, has had on the research group. In other words, the researcher aimed at determining what the project achieved. Summative evaluation can transpire during the project implementation, but often takes place at the culmination of a project (Evaluation Toolbox, 2010).

The table proposed by Evaluation Toolbox (2010) and based on Owen and Rogers (1999: 39-62), also covers summative education. This table has been adapted in Table 6.3, to indicate which data collection techniques for summative evaluation were used in this study.

Table 6.3: Summative Evaluation Data Collection Techniques

	Outcome
When	Project implementation and post-project
Why	"To assess whether the project has met its goals, whether there were any unintended consequences, what were the learnings, and how to improve" (Evaluation Toolbox, 2010)
Techniques used in this study	Semi-structured interviews

6.6 SUMMARY

This chapter gave an overview of the research design followed. The discussion started with an outline of the non-empirical part of this study, which consists of a literature study of the concepts, and an overview of the empirical part of the study, consisting of case studies. The concepts ‘literature study’ and ‘case study method’ were discussed, followed by a description of the various methods used in the case study, namely sampling method and triangulation, PAR, and prototyping. The focus of the discussion then shifted to the various data collection methods used, namely participant observation, interviews as well as testing and prototyping. This was followed by an overview of the research questions asked during the interviews, and finally, a description of the methods of analysis and of evaluation followed for this study.

The next chapter comprise the actual empirical part of this study.

CHAPTER 7
RESULTS: PRESENTATION AND DISCUSSION

7.1 INTRODUCTION

The analysis of findings in the two case studies was done through a process of formative and summative evaluation (see 6.5). The formative evaluation was applied through a process of participatory action research (PAR) (see 6.2.2.3), using notes taken during meetings, training sessions with the members of these case studies, as well as e-mail correspondence between the VRE design team (consisting of a VRE designer and the researcher of this study), and the members of these VRE groups (Case Study A and Case Study B), as data collection methods. The platforms (tools) that were identified as being suitable for a technological framework for a VRE in these case studies were designed through a process of testing and prototyping (see 6.2.2.4), and are summarised in Figures 7.4 and 7.5. The formative evaluation is followed by a process of summative evaluation consisting of semi-structured interviews (see 6.2.2.5) with the members in each of these case studies. The answers received are mapped to findings in the literature as well as results received through the formative evaluation process.

7.2 FORMATIVE EVALUATION

In Section 6.5 of this study, the researcher presented Evaluation Toolbox's (2010) table based on Owen and Rogers (1999: 39-62). This table (see Table 6.2) lists the various dimensions of formative evaluation. Each of these dimensions (see 6.5.4.1) - pro-active, clarificative, interactive, and monitoring - has specific techniques that can be used to gather information. These dimensions have been indicated in each of the steps in development process of the VRE technological frameworks, for each of the case studies. The techniques used to gather information for the evaluation of each of the two case studies consist of notes taken during meetings and e-mails between members of these case studies. For the sake of anonymity, these documents (notes and e-mails) were not listed in the bibliography, but have been numbered AD1, AD2, etc. for Case Study A, and BD1, BD2 etc. for Case Study B.

7.2.1 Case Study A

As mentioned in 6.2.2, Case Study A consisted of five postgraduate researchers, a promotor acting as VRE Manager, and a laboratory manager acting as VRE Champion, and who was also co-managing the VRE. This case was using natural science-oriented data and laboratory / experimental methods.

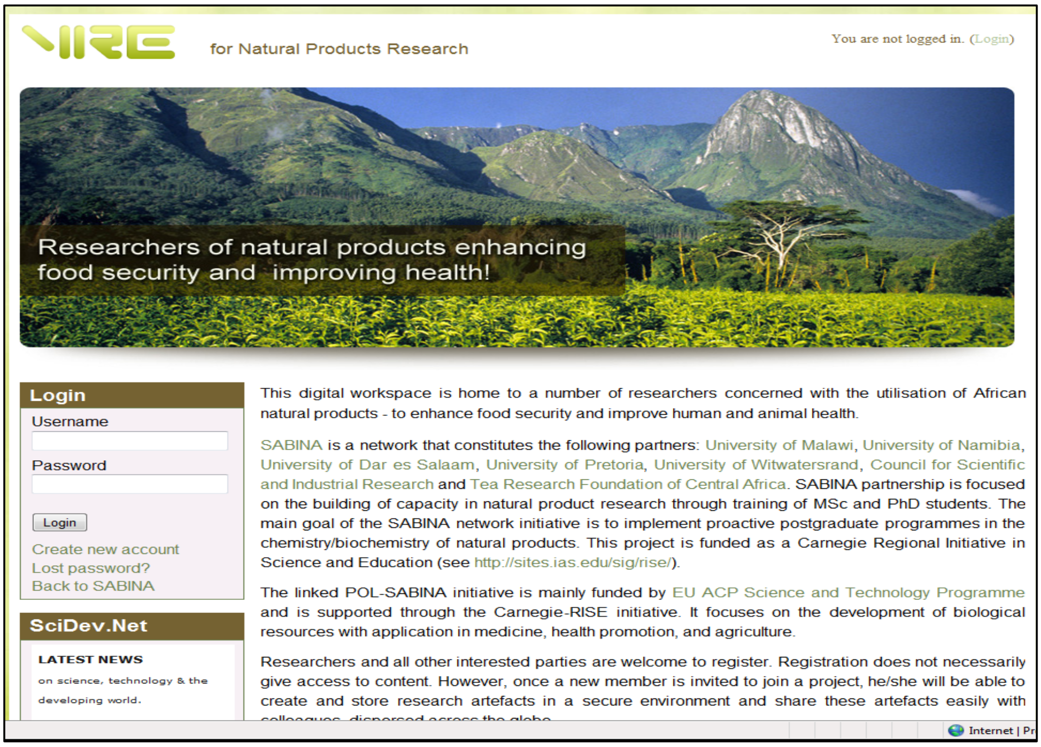
7.2.1.1 Identify Case Study A (Pro-Active Dimension)

The researcher of this study is a member of staff responsible for RDM practices at the University of Pretoria. Between October 2009 to March 2010, a survey was conducted by the Department of Library Services at the same University, to determine what is happening with regards to RDM at the University. Following this survey, the Director of Library Services compiled a project outline for the implementation of RDM at the University on 22 October 2012 (e-mail, 22 October 2012, AD1). This was then discussed with the Vice Principal Research, as well as with the Library Advisory Committee (a committee that was set up to advise the Library on strategic matters). The report was subsequently approved. In this report, a recommendation was made that the Library Services would initiate an RDM pilot project at the University. The idea was that the pilot project would give the involved staff members a good idea of how to take this project further. The researcher of this study, together with a member of the executive team of the Library Services, then identified a faculty that expressed an interest and concern about the RDM practices at the University. The Deputy Dean Research of this faculty was approached for the possibility of his faculty being a pilot study for RDM (e-mail, 15 January 2013, AD2). This was followed by a meeting on 11 April 2013 between members of the Library Services (this researcher and the manager of the institutional repository), the Deputy Dean Research of the particular Faculty and the Chair of the Ethics Committee of the Faculty, as well as the head of one of the institutes in the Faculty (e-mail, 15 April 2013, AD4). The outline of the discussion was already set up on 1 February 2013 (e-mail, 1 February 2013, AD3), and included matters such as: DMPs, metadata schemas, the types of data that they work with, the possibility of using the institutional repository of the University as a data repository, depositing data, whether the data should be open, data on paper, lab notebooks, electronic data, etc. (e-mail, 15 April 2013, AD4). During the meeting, it was emphasized that information gathered through the Ethics Committee should not be duplicated. A decision was then taken that one of the research groups that was represented in the meeting would be used as a pilot study (Case Study A in this thesis) (e-mail, 15 April 2013, AD4).

7.2.1.2 Explore The Case Study (Pro-Active Dimension)

The first contact session between the researcher of this study, a member of the Library Executive team and the Head of the research group of Case Study A (later also the VRE Manager of this group), as well as the Laboratory Manager (later the VRE Champion of this group) took place on 29 April 2013 (e-mail, 29 April 2013, AD5). During this meeting, it was found that there was only one dedicated computer workstation where each student had their own space to upload their data folders. The lab books were found to be in paper format. At that stage, it was speculated that the data could be managed by creating a space on an existing VRE platform, called the Natural Products VRE (a Southern African VRE running on Moodle - see Figure 7.1 - in which a number of Southern African institutions were involved) where this group of researchers could upload their day-to-day data (active data).

Figure 7.1: Natural Products VRE


VRE
for Natural Products Research
You are not logged in. (Login)
Researchers of natural products enhancing food security and improving health!
Login
Username
Password
Login
Create new account
Lost password?
Back to SABINA
SciDev.Net
LATEST NEWS
on science, technology & the developing world.
Internet | Pr

This digital workspace is home to a number of researchers concerned with the utilisation of African natural products - to enhance food security and improve human and animal health.

SABINA is a network that constitutes the following partners: University of Malawi, University of Namibia, University of Dar es Salaam, University of Pretoria, University of Witwatersrand, Council for Scientific and Industrial Research and Tea Research Foundation of Central Africa. SABINA partnership is focused on the building of capacity in natural product research through training of MSc and PhD students. The main goal of the SABINA network initiative is to implement proactive postgraduate programmes in the chemistry/biochemistry of natural products. This project is funded as a Carnegie Regional Initiative in Science and Education (see http://sites.ias.edu/sig/rise/).

The linked POL-SABINA initiative is mainly funded by EU ACP Science and Technology Programme and is supported through the Carnegie-RISE initiative. It focuses on the development of biological resources with application in medicine, health promotion, and agriculture.

Researchers and all other interested parties are welcome to register. Registration does not necessarily give access to content. However, once a new member is invited to join a project, he/she will be able to create and store research artefacts in a secure environment and share these artefacts easily with colleagues dispersed across the globe.

It was further foreseen that they would keep their paper lab books, but with better cross-references to the electronic data, and that their lab books would be digitised at the end of their study. To store their final (analysed) data, it was suggested that a closed space be created for it on the existing Institutional Repository of the University, which was running on DSpace (e-mail, 29 April, AD5).

7.2.1.3 Expand The Exploration Of The Case – Needs Identification For The VRE (Pro-Active Dimension)

On 10 May 2013, the researcher of this study, a member of the Library Executive team, a Library IT Specialist (later the VRE designer), and the student researchers involved in the project, had a second contact session (e-mail, 10 May 2013, AD6). The facilitator of the Natural Products VRE also attended the meeting. During this meeting, a demonstration was given on the Natural Products VRE (done on Moodle) by the facilitator of this VRE, and the students were asked to formulate, as a group, their needs in terms of RDM. The plan was to discuss these needs during a follow-up meeting. In order to demonstrate how it would work, it was further decided to upload the data of one of the students that had completed her studies, onto Moodle and on an instance of DSpace.

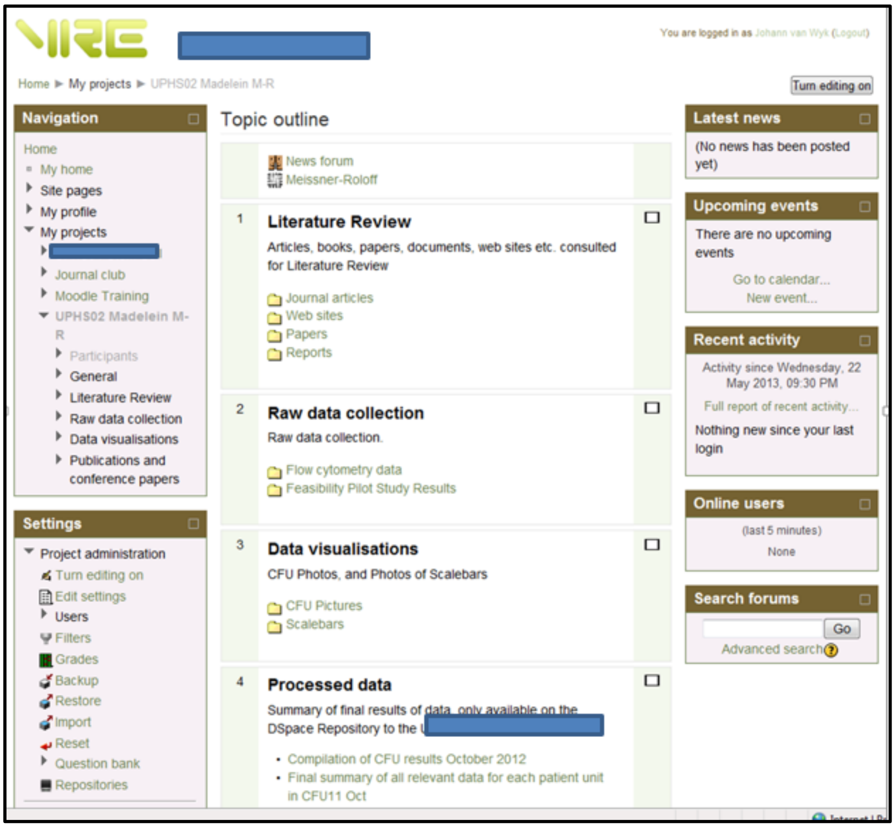
Following the meeting, a document outlining the pilot project process was compiled by the member of the Library Executive Team (Document, 16 May, AD8). This document indicated that a bottom-up approach would be better, as researchers are primarily responsible for the management of their own data. The document further mentioned that the pilot project would be hosted on a specific server on the Hatfield Campus of the University, but emphasized that the group might need their own server in the medium term to ensure security of their data. The document also indicated that in the longer term, data could be harvested from the project and archived and curated in a yet-to-be-identified system at the University.

A site was created for the group on 16 May 2013 on Moodle (e-mail, 16 May 2013, AD7). The data of the student that had completed her studies were then uploaded onto the site on 17 May 2013 (e-mail, 17 May 2013, AD9). These included Flow Cytometry data with fcs data extensions, which had been generated with the Kaluza programme, Excel files, and Jpeg files, as well as pdf copies of articles consulted in the study. The hard copy lab book of this student was also digitised and then uploaded onto Moodle. The thesis of the student was uploaded on an instance of DSpace, and linked to the Moodle site (e-mail, 17 May 2013, AD9).

7.2.1.4 Demonstration Of The Initial VRE Prototype (Pro-Active Dimension)

A third contact session was scheduled on 19 July 2013 between the researcher of this study, a member of the Library Executive team, a Library IT Specialist (later the VRE designer), and the Head of the project (later the A-VRE-M), the laboratory manager (later the A-VRE-C) and the student researchers involved in the project (e-mail, 11 July 2011, AD10). During this meeting, the uploaded files of the student that had completed her studies were demonstrated on a prototype created on Moodle and linked to DSpace (see Figure 7.2). An effort was made to structure the folders in such a manner that it would support the Research Lifecycle (see 3.4.1 and Figure 3.4). This was then followed by a discussion on the needs of the group.

Figure 7.2: Prototype On Moodle



VRE [redacted]

You are logged in as Johann van Wyk (Logout)

Home ▶ My projects ▶ UPHS02 Madelein M-R

Turn editing on

Navigation

- Home
- My home
- Site pages
- My profile
- My projects
  - [redacted]
  - Journal club
  - Moodle Training
  - UPHS02 Madelein M-R
    - Participants
    - General
    - Literature Review
    - Raw data collection
    - Data visualisations
    - Publications and conference papers

Settings

- Project administration
  - Turn editing on
  - Edit settings
  - Users
  - Filters
  - Grades
  - Backup
  - Restore
  - Import
  - Reset
  - Question bank
  - Repositories

Topic outline

News forum
Meissner-Roloff

1 **Literature Review**
Articles, books, papers, documents, web sites etc. consulted for Literature Review
Journal articles
Web sites
Papers
Reports

2 **Raw data collection**
Raw data collection.
Flow cytometry data
Feasibility Pilot Study Results

3 **Data visualisations**
CFU Photos, and Photos of Scalebars
CFU Pictures
Scalebars

4 **Processed data**
Summary of final results of data, only available on the DSpace Repository to the [redacted]
- Compilation of CFU results October 2012
- Final summary of all relevant data for each patient unit in CFU11 Oct

Latest news
(No news has been posted yet)

Upcoming events
There are no upcoming events
Go to calendar...
New event...

Recent activity
Activity since Wednesday, 22 May 2013, 09:30 PM
Full report of recent activity...
Nothing new since your last login

Online users
(last 5 minutes)
None

Search forums
Go
Advanced search

7.2.1.5 First Formative Evaluation: Adapt The VRE (Interactive Dimension)

a) Issues / Commentary That Led To Adaptations

During the meeting on 19 July 2013, the student researchers in Case Study A indicated that they did not need many additional features or trimmings such as social media features; however, they expressed a need for a place to back-up their data, and would prefer to be able to synchronise (known in the vernacular as sync) their data on Google Drive with the Moodle instance. The wish was also expressed that in future, it would be great to be able to access their processing and analysis programmes within the VRE (notes, 19 July 2013, AD11).

b) Adaptations

Following the expressed needs of the student researchers, a decision was made to adapt Moodle as a VRE site for the group, and to register everyone on it so they could play around with it and test it (notes, 19 July 2013, AD11). The VRE Designer subsequently registered the members of the group on 29 July 2013, and added a link to Google Drive for synchronising purposes. A server was also installed at the location where Case Study A is situated (e-mail, 29 July 2013, AD12). The VRE Designer (VRE-D) confirmed the registration of members in an e-mail sent to all the members of the group: “I finished the registration of everyone involved and will send the information to each person individually. I am just finishing one or two details - the Google Drive link specifically for syncing purposes, before doing so” (e-mail, 29 July 2013, AD12).

c) Qualitative Commentary From VRE Members To Confirm The Changes / Work Done On The VRE

In the document that outlined the pilot project process (Document, 16 May 2013, AD8), a member of the Library Executive Team mentioned the following: “for the pilot we will make use of the VRE server at FABI, but it could be that this research group will in future need their own server, especially for security reasons.” As a follow-up to this, the VRE-D commented the following in an e-mail dated 29 July 2013: “I am still awaiting a quotation from Dell regarding a server” (e-mail, 29 July 2013, AD12). The installation of a server at the campus where Case Study A is located was then confirmed in a meeting held on 1 October 2013 between the VRE-D, the researcher of this study, and the A-VRE-C. The VRE-D stated: “there is a server at this campus and a backup server at the Merensky Library on the Hatfield Campus” (notes, 1 October 2013, AD15).

7.2.1.6 Training Session 1 (Clarificative Dimension)

On 5 August 2013, the VRE-D and the A-VRE-C, came to an agreement that it would be better to do group training for members of Case Study A, and to do more extensive training for the A-VRE-C, as she would be expected to assist members of the group (e-mail, 5 August 2013, AD12). The result of this decision then led to a training session on the Moodle platform with the A-VRE-C on 19 September 2013, which was conducted by VRE-D as well as the researcher of this study (e-mail, 12 September 2013, AD13) (see also the VRE-D’s answer to Question 43).

7.2.1.7 Identify More Needs And Adaptations To The VRE (Clarificative Dimension)

A follow-up meeting was held on 1 October 2013 between the researcher of this study, the VRE-D and the A-VRE-C (e-mail, 1 October 2013, AD14). During this meeting, the A-VRE-C requested that the file sizes that the system could handle, should be increased, and that the roles and rights of the members should be clearly defined. It was decided that the supervisor / promotor would have VRE Manager rights and that the Laboratory Manager would act as VRE Champion, but also have VRE Manager rights. The student researchers would only have rights to access and edit their own spaces and to read and access shared spaces (notes, 1 October 2013, AD15). The VRE-D followed up on this in an e-mail sent on 1 October 2013: “The role assignment names have been changed (in order of rights) Manager > Supervisor > Non-editing Supervisor > Student” and “the file size upload has been increased to 5GB (5120mb)” (e-mail, 1 October 2013, AD14). Other issues that were discussed were file-naming conventions for the files that are uploaded, problems with the University’s Firewall that was blocking the synchronising of files from Google Drive, as well as the issue of versioning.

7.2.1.8 Second Formative Evaluation: Replace The Moodle VRE Platform With An Instance On Alfresco (Interactive Dimension)

(a) Issues / Commentary That Led To Adaptations

During the discussion on 1 October 2013 (notes, 1 October 2013, AD15), it was found that members of the group needed a versioning function, but Moodle, however, could not fulfil this need. Two possibilities were considered: either integrate Moodle with Alfresco to provide a versioning function for Moodle, or replace the instance on Moodle with an instance on Alfresco. The integration with Alfresco, however, proved to be problematic, as can be seen in the VRE-D's comment in an e-mail on 1 October 2013 (e-mail, 1 October 2014, AD14): "The Alfresco integration is still one way (the development environment keep [sic] crashing), but I am busy with it."

(b) Adaptations

During the meeting on 1 October 2013, it was decided that replacing the Moodle instance with an instance on Alfresco would be a better option. Alfresco, a document management system, was seen as a more user-friendly system. The Alfresco VRE also had the following positive points:

- It had an efficient versioning function (versions of previous documents are kept and stored);
- It had a very good metadata function that can help one to find documents again;
- It could easily be integrated with other software;
- It gave the promotor / supervisor an overview of his / her students' progress;
- It had a very good workflow management system (valuable for moderation, peer review);
- It had very good rights management (determining level of access);
- It could synchronise with file management software such as Google Drive or Dropbox (a real need that the researchers identified);
- One could also easily drag and drop files from a hard drive / flash disk into Alfresco;
- It enabled users to do file sharing; and

- It had a mobile application (app), which is valuable for researchers (they could use a mobile device to upload files, images, video etc. whenever they are busy in the lab, or interviewing subjects) (notes, 1 October 2013, AD15).

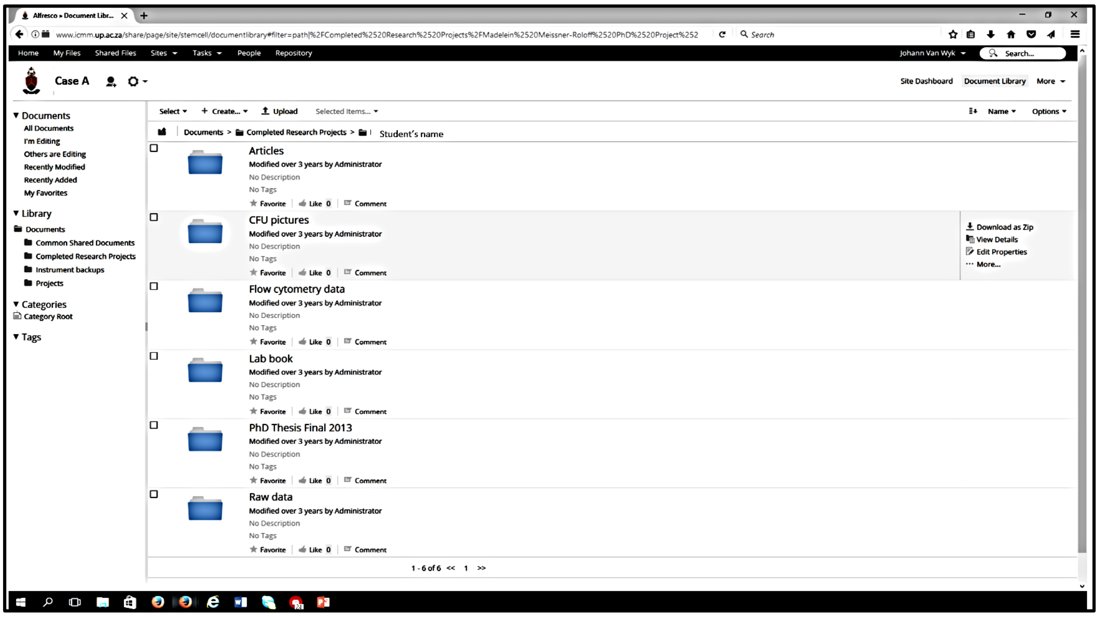
The group's site (with members' profiles, their files and data) was subsequently migrated by the VRE-D from Moodle to Alfresco (see an example of a researcher's page on Alfresco in Figure 7.3).

(c) Qualitative Commentary From VRE members / VRE-D To Confirm The Value Of, And Changes / Work Done On The VRE.

The VRE-D mentioned that Alfresco "has a strong versioning function, in other words, the system could keep multiple versions of a data file, and the system has a very good metadata function" (notes, 1 October 2013, AD15). Continuing, the VRE-D mentioned that "it also has a good workflow system and synchronising function, as well as access to social tools and functions" (notes, 1 October 2013, AD15).

The migration to Alfresco was completed by 21 October 2013, confirmed through this comment in an e-mail from the VRE-D to the A-VRE-C and the researcher of this study: "The URL to the new VRE is http://icarus.up.ac.za:8080/share" (e-mail, 21 October 2013, AD23).

Figure 7.3: Example Of A Researcher's Page On The Alfresco VRE Instance


Alfresco » Document Libr...
www.icmm.up.ac.za/share/page/site/stemcell/documentlibrary#filter=path|%2FCompleted%2520Research%2520Projects%2FMadelein%2520Meissner-Roloff%2520PhD%2520Project%252
Search
Home
My Files
Shared Files
Sites
Tasks
People
Repository
Johann Van Wyk
Search...
Case A
Site Dashboard
Document Library
More
Documents
All Documents
I'm Editing
Others are Editing
Recently Modified
Recently Added
My Favorites
Library
Documents
Common Shared Documents
Completed Research Projects
Instrument backups
Projects
Categories
Category Root
Tags
Select
Create...
Upload
Selected Items...
Name
Options
Documents > Completed Research Projects > Student's name
Articles
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
CFU pictures
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
Download as Zip
View Details
Edit Properties
More...
Flow cytometry data
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
Lab book
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
PhD Thesis Final 2013
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
Raw data
Modified over 3 years by Administrator
No Description
No Tags
Favorite | Like 0 | Comment
1 - 6 of 6 << 1 >>

7.2.1.9 Training Session 2 (Clarificative Dimension)

The migration to Alfresco necessitated another training session, but this time all the student researchers as well as the VRE Manager and the VRE Champion of the group were included. This session was conducted by the VRE-D on 6 December 2013. The VRE Manager of Case Study B also attended this training session. The session consisted of an overview of the VRE system and a hands-on training session on the system in the Library Training Laboratory. This training was confirmed by the VRE-D in an e-mail to the researcher of this study and a member of the Library Executive on 9 December 2013: “I completed the training sessions with the A-VRE-M [name anonymised] and his students last Friday” and the B-VRE-M [name anonymised] was also present and she likes what we have done, and is definitely interested in such a system” (e-mail, 9 December 2013, AD16) (see also the VRE-D’s answer in Question 43).

7.2.1.10 Identify Additional Needs For Adaptations To The VRE (Clarificative Dimension)

The members of Case Study A reported a number of teething problems in the early stages of using the VRE platform (Alfresco). On 11 March 2014, the A-VRE-C reported a problem with the updating of files and synchronising of files on the system, to the VRE-D (e-mail, 11 March 2014, AD17) (see also answer to Question 27). On 8 July 2014, a problem with the saving of Google Docs onto Alfresco was reported, which turned out to be a problem with Google’s code constantly changing (e-mail, 8 July 2014, AD18). On 8 August 2014, the A-VRE-C reported that two of the student researchers (A-R2 and A-R5) were experiencing problems - their folders were being duplicated by the system (e-mail, 8 August 2014, AD 20). The system was offline on 15 September 2014 because of a network problem, but was fixed on the same day (AD, 15 September 2014, AD21).

7.2.1.11 Third Formative Evaluation: Adapt The VRE (Interactive Dimension)

(a) Issues / Commentary That Led To Adaptations

In 7.2.1.10, a number of teething problems were raised, which led to small adaptations / fixes to the VRE. These can be confirmed through the following commentary in e-mails between the A-VRE-C and the VRE-D:

- A-VRE-C: “Several people are currently trying to use the VRE and they can login but cannot do anything (change passwords, update, sync, etc.). The error we are getting is 500 internal server error. Can you please urgently have a look?” (e-mail, 11 March 2014, AD17).
- A-VRE-C: A student also mentioned that “the syncing is not working properly” (e-mail, 11 March 2014, AD17).
- A-VRE-C: “I am having a problem saving a document I edited online using Google Docs. Can you maybe check this out if you have time?” (e-mail, 8 July 2014, AD18).
- A-VRE-C: “Could you please do me a favour and check out A-R4’s [name anonymised] folder in the projects folder? There is a folder under his name called projects, where all the folders under A-R4 [name anonymised] are being duplicated, some folders have items in them and others are empty. This seems to be happening in A-R2’s [name anonymised] folder as well” (8 August 2014, AD20).
- The VRE-D sent an e-mail to the A-VRE-C on 15 September 2014, and asked the following: “I am unable to see the server? Seems like all the machines in the building is [sic] off-line. Can you confirm?” The VRE-D responded via e-mail: “Yes, we are having major problems with the Internet this side. I cannot get on. I get service temporarily unavailable” (15 September 2014, AD21).

(b) Adaptations

The adaptations in the third formative evaluation were more focused on network, server, and external problems. The problems that were encountered with the 500 internal server error message were caused by a routing problem on the server where the VRE was located, and was speedily solved by the VRE-D (e-mail, 11 March 2014, AD17).

The VRE-D reported the problem with Google Docs to the Alfresco Engineers for a solution (e-mail, 8 July 2014, AD18). He also moved the data that were duplicated back to their original folders (e-mail, 8 August 2014, AD20). The damage caused to the server by a network downtime was fixed by the VRE-D and the VRE operated fine after that (e-mail, 15 September 2014, AD21).

(c) Qualitative Commentary From VRE members / VRE-D To Confirm The Changes / Work Done On The VRE

- The VRE-D confirmed in an e-mail that the 500 internal server error had been addressed. His comment was: "There was a routing problem, it's sorted" (e-mail, 11 March 2014, AD17).
- The VRE-D's reply to the problem with the saving of a document that was edited online with Google Docs, was: "The problem seems to be with the Google Code changing the whole time, as you can see at the bottom of this forum https://forums.alfresco.com/forum/installation-upgrades-configuration-integration/installation-upgrades/google-docs-integratio-0. The Alfresco Engineers are looking into it" (e-mail, 8 July 2014, AD18).
- The VRE-D also solved the issue of the duplication of A-R2 and A-R4's folders, and commented on this as follows: "Ok I looked at the data and moved it back… it says that the folders were changed 5 hours ago. I need to see what caused it" (e-mail, 8 August 2014, AD20).
- The VRE-D responded to the problem that was reported with regards to the network that was down, and commented the following: "I see there is an emergency change for the network, will see what is going on and keep you posted." He came back and commented the following: "The server did suffer a bit because of the network down time. It is, however, up and running now and all seems to be fine" (e-mail, 15 September 2014, AD21).

7.2.1.12 Implement Changes To The VRE (Monitoring Dimension)

All the changes and issues mentioned in 7.2.1.11 had been solved and implemented by the VRE-D by 15 September 2014.

7.2.1.13 Identify Further Needs And Adaptations (Clarificative Dimension)

On 20 May 2015, the VRE-D and the researcher of this study had a meeting with the A-VRE-C to determine further hardware and software needs with regards to the VRE (e-mail, 20 May 2015, AD22).

7.2.1.14 Fourth Formative Evaluation: Adapt The VRE (Interactive Dimension)

(a) Issues / Commentary That Led To Adaptations

The following decisions were taken during the meeting that was held on 20 May 2015 between the researcher of this study, the VRE-D and the A-VRE-C, that the VRE D would:

- Install a 15 TB NAS (Network Attached Storage) device at the facility where the case study is operating in;
- Back-up the Facsaria, Galios and Affymetrix machine data to the NAS;
- Replicate the NAS at the laboratory to the one in the Library on the Hatfield Campus of the University;
- Arrange for the installation of a network point to connect the Galios machine to the network;
- Investigate the possibility of upgrading the Affymetrix machine to Windows 7;
- Install a new server room with 24-hour air-conditioning, four network points, two UPS's (uninterrupted power supply), and access control;
- Address a firewall issue with the Affymetrix machine;
- Provide one of the student researchers' access details to a virtual machine environment to test some of the software and processing power (e-mail, 20 May 2015, AD22).

(b) Adaptations

The issues mentioned in 7.2.1.14 a) were addressed by the VRE-D in the following manner:

- One of the 15 TB NAS (Network Attached Storage) devices that was situated at the Hatfield Campus, was moved to the facility where the case study was operating in;
- The Facsaria, Galios and Affymetrix machine data were backed-up to the NAS device;
- The NAS device at the laboratory of the institution where the case study operated, was set up in such a manner that it could be replicated to a NAS device in the Library on the Hatfield Campus of the University;
- The firewall issue with the Affymetrix machine was resolved; and
- One of the student researchers (A-R2) was provided access to a virtual machine environment to test some of the software and processing power.

The A-VRE-M arranged for the installation of a network point to connect the Galios machine to the campus network, and also identified a potential server room in the facility where the case study was operating in. The identified room was then equipped as a server room.

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE.

Some of the adaptations mentioned was communicated orally by the VRE-D to the researcher of this study. The A-VRE-M, however, confirmed the following in a reply to an e-mail:

- "I am delighted to inform you that we have managed to identify a room near my office which we can dedicate entirely as a server room. Please would you let me know when you will next be on this Campus so that I can show you and so that we can effect the necessary alterations as specified […]";
- "I will arrange to have a network point installed near the Galios machine" (e-mail, 24 May 2015, AD22).

7.2.1.15 Implement Changes To The VRE (Monitoring Dimension)

The needs identified in 7.2.1.14 were addressed by the VRE-D over a period of 4 weeks after 20 May 2015, with assistance from the University's IT department.

7.2.2 Case Study B

In 6.2.2, it was mentioned that Case Study B consisted of four postgraduate researchers, a promotor acting as VRE Manager, as well as a librarian. The same VRE designer was used as in Case Study A. This case study used human-oriented data and survey instruments as data collection method.

7.2.2.1 Identify A New Group That Would Form Case Study B (Pro-Active Dimension)

On 26 November 2013, the researcher of this study received an e-mail from the promotor / supervisor of a research group that was working with human-oriented data (e-mail, 26 November 2013, BD1). The promotor had heard of the VRE project of Case Study A and got permission from the Deputy Dean Research of her faculty to send an e-mail requesting for the creation of a second VRE for this particular field and group of postgraduate researchers working in it. In her request, the promotor indicated that they were urgently in need of a system that could help them manage their data in a more structured way, in order to enable others (e.g. publishers) to interrogate the data, after these projects had been completed. This group has been called Case Study B in this study, and the promotor was identified as the contact person and VRE Manager for this group, and named B-VRE-M further in this study.

7.2.2.2 Training Session 1 (Pro-active Dimension)

The B-VRE-M for Case Study B was subsequently invited to attend the hands-on training session (Training Session 2) for Case Study A that was held on Friday 6 December 2013, in order to acquaint herself with the system and gain an idea whether the system would address the group's needs (e-mail, 26 November 2013, BD2). She then attended the training session on 6 December 2013 (e-mail, 13 January 2014, BD3) (see also the VRE-D’s answer to Question 43). This training session became training Session 1 in Case Study B.

7.2.2.3 Explore The Case And Identify Needs (Clarificative Dimension)

The researcher of this study contacted the B-VRE-M on 13 January 2014 to arrange a meeting with her and the VRE-D, to discuss the needs and specifications she and her group of postgraduate researchers would have for a system that can assist with the management of their research data (e-mail, 13 January 2014, BD4). A meeting was subsequently held on 17 January 2014, which was attended by the B-VRE-M, the VRE-D, a colleague of the B-VRE-M (who would also be a postgraduate researcher in the group), as well as the researcher of this study. Alfresco was identified as a system that would be able to meet all their needs.

7.2.2.4 First Formative Evaluation: Create The First Instance Of The VRE

(a) Issues / Commentary That Led To Adaptations

During the meeting on 17 January 2014, it was found that the group had the following needs with regards to a VRE:

- A big need for a workflow function between the promotor / supervisor (acting as VRE Manager) and the student researchers;
- A versioning function;
- A place to archive their data; and
- A survey tool in the system (notes, 17 January 2014, BD5).

Alfresco could provide in most of these needs, but the VRE-D indicated that he would have to plug a survey tool into the system. However, this would not be problematic (notes, 17 January 2014, BD5).

(b) Adaptations

After the meeting on 17 January 2014, the VRE-D created an instance of a VRE on Alfresco that accommodated the needs expressed during the meeting (e-mail, 22

January 2014, BD7)**.** The B-VRE-M also sent through a list of the student researchers that should be registered on the system and should be trained in the system (e-mail, 17 January 2014, BD6).

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE:

- The B-VRE-M confirmed the meeting that was held in an e-mail: "Thank you that you visited us to investigate our need for a data management system" (e-mail, 17 January 2014, BD6);
- The B-VRE-M listed the students that would be involved (e-mail, 17 January 2014, BD6); and
- The VRE-D stated in an e-mail: "I have just finished with the creation of a site, and is just waiting on the network personnel to open it on the 'firewall'. As soon as it is open I will send more detail on where the students can visit it, so that they can get a feel for the site" (e-mail, 22 January 2014, BD7).

7.2.2.5 Training Session 2 (Clarificative Dimension)

A hands-on training session on the various functionalities of the Alfresco platform was arranged on 27 January 2014, for everyone involved in Case Study B, on the campus where their project was centred (e-mail, 22 January 2014, BD7). The participants included the B-VRE-M, all the postgraduate students in the group, and the librarian involved in this field (e-mail, 28 January 2014, BD8). The same VRE-D responsible for Case Study A conducted the training session (e-mail, 28 January 2014, BD8). The researcher of this study also gave a short overview on File Naming Conventions, which could be of value to the members of the group when they organise their files and folders (see also the VRE-D's answer to Question 43). The VRE-D followed this training session up with an e-mail where he repeated the main points that were touched on during the training session on 27 January 2014: "The web-address of the system is http://icarus.up.ac.za:8080/share. The username and password are the e-mail address and then the password as provided during the session. Remember the first screen is a person's Personal Dashboard and from there one has to go the group's site. All documents are available via the 'Document Library' on the top right of the screen. 'Site

Supervisors' have access to all the data and the rest have access to their own folders and guides. Every document and guide has extra 'document / Folder Actions on the right side after it has been selected. The most important of these are 'Upload new version' and also 'Start Workflow'. The Apple version of the Application (app in the vernacular) is available via the Apple Store and for Android via the Google Playstore" (e-mail, 28 January 2014, BD8).

7.2.2.6 Implement The First Instance Of The VRE

After the training session on 27 January 2014, the VRE-D encouraged members of the VRE to take their time and work through the system and 'play' around with it (e-mail, 28 January 2014, BD8). He also gave details regarding information needed to use an application for mobile devices (e-mail, 28 January 2014, BD8).

7.2.2.7 Expand The Exploration Of Needs, And Identify New Needs And Adaptations To The VRE (Interactive Dimension)

On 28 January 2014, the B-VRE-M reported that the system was not sending out e-mail notifications when someone had uploaded something on the VRE (see answer to Question 27 under 7.3.1.1). She was also unable to log onto the system (e-mail, 28 January 2014, BD9). The VRE-D speedily resolved these problems.

7.2.2.8 Second Formative Evaluation: Adapt The VRE To Meet The Needs

(a) Issues / Commentary That Led To Adaptations:

The B-VRE-M reported the following in an e-mail to the VRE-D on 28 January 2014:

- "B-R4 [name anonymised] mentioned that she has uploaded a number of things, but I have not yet received an e-mail notification that there is something I need to look at" (e-mail, 28 January 2014, BD9);
- "I tried to log in but only get a 'system error' message" (e-mail, 28 January 2014, BD9);
- "With the iPad app the security settings are not the same as with the laptop – the B-R4 [name anonymised] could see all the data from Case Study A, and could

download and change these – all the students' data are open on the system" (e-mail, 28 January 2014, BD9).

(b) Adaptations

The VRE-D followed up on the B-VRE-M's e-mail message sent out on 28 January 2014 and corrected a problem with the workflow of the system that was blocking the system from sending out e-mail notifications (e-mail, 28 January 2014, BD9). He also sorted out the login problem. The VRE-D furthermore corrected the problem with security settings that exposed data of students from Case Study A.

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE:

- The VRE-D stated in an e-mail on 28 January 2014: "it looks like the workflow is now fixed" (e-mail, 28 January 2014, BD9);
- In response to the corrections made with regards to the security settings that exposed data of students from Case Study A, the VRE-D stated the following: "I have looked at the problem, but unfortunately it was my fault… it was Case Study A's replication data that had 'Site-Contributors' rights. Thank you for bringing this to my attention, I apologise" (e-mail, 28 January 2014, BD9).

7.2.2.9 Implement Changes To The VRE

The corrections / changes made to the issues mentioned in 7.2.2.8 were effected by the VRE-D on the same day it was reported to him, namely 28 January 2014.

7.2.2.10 Training Session 3 (Clarificative Dimension)

The B-VRE-M requested a training session on 17 February 2014 for two of the student researchers (B-R1 and B-R2), who stayed geographically far away from the campus (e-mail, 14 February 2014, BD10). The one respondent stayed in another country, and the other, in another province. They also needed a survey tool to do their research, as part of the VRE, which was one of the needs that were mentioned in 7.2.2.4. On 17 February

2014, one of the student researchers (B-R2) as well as the B-VRE-M were trained in the survey tool and the way it operates in the VRE. The other researcher (B-R1) was only trained at the end of March 2014 (e-mail, 13 March 2014, BD11) (see also the VRE-D's answer to Question 43).

7.2.2.11 Identify Further Needs Or Adaptations (Monitoring Dimension)

The VRE system operated without any problems for the rest of 2014 after the implementation of changes on 28 January 2014. The VRE-D and the researcher of this study then arranged a follow-up meeting with the B-VRE-M in early 2015, which took place on 19 January 2015 (notes, 19 January 2015, BD12). During this meeting, the B-VRE-M expressed her need to be able to upload a video or audio file via an iPad to the VRE platform. The VRE-D explained to her that it is possible to upload a video, audio or image file through the Alfresco App, or alternatively, take a video, audio and images directly from the Alfresco platform using the device's camera and microphone (notes, 19 January 2015, BD12). The B-VRE-M also stated her need for a calendar function for everyone, and some or other mechanism to be able to monitor her student researchers' progress (notes, 19 January 2015, BD12). The VRE-D explained to her how the calendar function works, and how she can add a calendar 'dashlet' onto the VRE platform (notes, 19 January 2015, BD12). The VRE-D also explained to her in detail how the workflow function would work between her as supervisor and the student researchers (notes, 19 January 2015, BD12).

7.2.2.12 Third Formative Evaluation: Adapt The VRE

(a) Issues / Commentary That Led To Adaptations

During the meeting held on 19 January 2015 the A-VRE-M mentioned:

- "I would like to upload a video or audio file from an iPad to Alfresco, how can I do this?" (notes, 19 January 2015, BD12); and
- "I would like to have a calendar feature for everyone, and would like to create a workflow for each person" (notes, 19 January 2015, BD12).

The VRE-D commented the following during the meeting held on 19 January 2015:

- "One can upload video and / or audio via the Alfresco app";
- "One can also take video, audio and images directly from the Alfresco app. It accesses the device's camera and microphone" (notes, 19 January 2015, BD12).

(b) Adaptations

The VRE-D explained and demonstrated how to upload a video and / or audio file via the Alfresco application. He also helped the B-VRE-M to download an Alfresco application (app) onto her laptop, and showed her how to take a video, record audio, and take photos directly through the Alfresco application, by using the device's camera and microphone (notes, 19 January 2015, BD12). The VRE-D also added a site calendar dashlet to the Alfresco site (notes, 19 January 2015, BD12).

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE

The changes done were minimal and consisted of the downloading of an Alfresco application on B-VRE-M's laptop and the adding of a dashlet for a site calendar on the VRE site. These were all done during the meeting held on 19 January 2015. The only commentary received was a word of thanks from the B-VRE-M (notes, 18 January 2015, BD12).

7.2.2.13 Implement Changes To The VRE

The changes mentioned in 7.2.2.12 was implemented during the meeting held on 19 January 2015.

7.2.2.14 Identify Further Needs Or Adaptations (Monitoring Dimension)

On 28 January 2015, the B-VRE-M indicated to the VRE-D that she had a problem accessing the VRE and that the system was not sending notifications of documents and files that had been uploaded on the system by her student researchers (e-mail, 28 January 2015, BD13).

7.2.2.15 Fourth Formative Evaluation: Adapt The VRE

(a) Issues / Commentary That Led To Adaptations

- On 28 January 2015, the B-VRE commented in an e-mail: “I cannot get into Alfresco and no-one gets messages indicating that there is a message for them – I am talking about the docs / videos that we uploaded” (e-mail, 28 January 2015, BD13);
- The VRE-D responded on 28 January 2015: “I can get into the system from my side. What is the error message that you receive from the system, or is nothing happening?” (e-mail, 28 January 2015, BD13). The B-VRE-M responded: “[…] thank you that you are looking into it – nothing is happening on my side.”
- On 30 January 2015, the B-VRE-M reported: “I could get onto Alfresco this morning, and were able to upload documents” (e-mail, 30 January 2015, BD14). She further mentioned: “[…] the problem is that no one, not even I, get notices indicating that I have uploaded documents for them” (e-mail, 30 January 2015, BD14). She continued: “[…] the videos that we uploaded for” B-R2 [name anonymised] “does not show on my site or her site on Alfresco” (e-mail, 30 January 2015, BD14).
- On 16 February 2015, the B-VRE-M reported the following: “[…] our problems on Alfresco has still not been solved – my students cannot access Alfresco, and we do not get feedback that something has been uploaded onto Alfresco – there is definitely something wrong with the communication […] we urgently need the communication and workflow function” (e-mail, 16 February 2015, BD15).

(b) Adaptations

The VRE-D investigated the problem with access and notifications and executed an upgrade of the system hoping that this would solve the problem (e-mail, 30 January 2015, BD 14). After the B-VRE-M mentioned that the problem with access and notification was persisting, the VRE-D investigated the problem further. He found that there was a problem with the notification function itself on the Alfresco system, and repaired this (e-mail, 16 February 2015, BD 15).

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE

- The first action by the VRE-D to repair the problems with access and notification can be confirmed through an e-mail sent by the VRE-D: “There seems to be a fault with the messaging system. I will upgrade everything over the weekend, and believe this will solve the problem” (e-mail, 30 January 2015, BD14).
- The second action by the VRE-D to repair the problem with the notification can be confirmed through a reply to an e-mail sent to the B-VRE-M on 16 February 2015: “I have repaired the notification on the system. Every person should now be receiving a notification when changes occur in their folders, as well as the group if there are changes in the general folder. E-mails should have come through this morning” (e-mail, 16 February 2015, BD15).
- The B-VRE-M confirmed this correction on the system on 16 February 2015 in an e-mail: “Thank you. Yes, I have received the e-mails – and B-R2 [name anonymised] in Ghana also informed me that she received the Alfresco-e-mail, and she was able to enter the link (e-mail, 16 February 2015, BD15).

7.2.2.16 Implement Changes To The VRE

All the corrections and changes mentioned in 7.2.2.15 were completed by 16 February 2015.

7.2.2.17 Identify Further Needs Or Adaptations (Monitoring Dimension)

The B-VRE-M requested a meeting on 20 May 2015 between her and the VRE-D to discuss the possibility of giving other co-supervisors (from another university) and other staff members in the department, access to the VRE, in order to monitor the student researchers’ progress (e-mail, 19 May 2015, BD16).

7.2.2.18 Fifth Formative Evaluation: Adapt The VRE

(a) Issues / Commentary That Led To Adaptations

The B-VRE-M sent an e-mail to the VRE-D on 19 May 2015 consisting of the following: “I would like to make an appointment with you to discuss the Alfresco page of my students with you. The manner in which I will be appointed from now on will make it crucial that all of them report back regularly about certain matters. I will also need to give others access to monitor the progress” (e-mail, 19 May 2015, BD16). By 25 May 2015, however, the B-VRE-M commented in an e-mail: “Today I discussed the basic principles of data management and workflow on Alfresco with the head of my department [name anonymised]. Could you perhaps also show her how to monitor the workflow? I also noticed that the co-study leader [name omitted for anonymity reasons] of B-R4 [name anonymised] and the co-study leader [name omitted for anonymity reasons] of B-R1 [name anonymised] does not appear on the lists of users yet” (e-mail, 25 May 2015, BD17).

(b) Adaptations

The VRE-D then demonstrated the workflow to the head of B-VRE-M’s department on 25 May 2015. Later that day, he also registered the co-study leaders (external from other universities) on the Alfresco system, with rights to monitor specific student researchers’ progress.

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE

No qualitative commentary was received, but there had also been no further complaints or requests regarding the provision of access to the Alfresco system to other external parties either.

7.2.2.19 Implement Changes To The VRE

As mentioned in 7.2.2.18, the VRE-D adapted the system to accommodate the registration of external users. This went into effect on 25 May 2015.

7.2.2.20 Identify Further Needs Or Adaptations (Monitoring Dimension)

The need for the activation of the survey tool on Alfresco was identified by the B-VRE-M on 1 November 2015.

7.2.2.21 Sixth Formative Evaluation: Adapt The VRE

(a) Issues / Commentary That Led To Adaptations

On 1 November 2015, the B-VRE-M sent an e-mail to the VRE-D in which she touched on the activation of the survey tool (called a questionnaire tool by her) on Alfresco. She stated: “The B-R2 [name anonymised], one of my PhD candidates, prefers to use the questionnaire tool on Alfresco which you demonstrated to the group. Could you please make this available on the Alfresco website, together with the procedure on how to use it on Alfresco? She would like to compile a questionnaire for the parents of premature babies. The respondents will be spread out all over the country, and will have to able to complete the questionnaire online. If I can remember correctly one could process the results on Alfresco?” (e-mail, 1 November 2015, BD18).

In reply to the e-mail sent by the B-VRE-M on 1 November 2015, the VRE-D stated in an e-mail: “Is it possible to set a date and time so that I can give training in the tool? It is a relatively easy system if there is no time for training. The system meets all the requirements that [name anonymised] mentioned” (e-mail, 4 November 2015, BD19).

(b) Adaptations

The VRE-D activated the survey tool on 4 November 2015 on Alfresco, by plugging it into the system.

(c) Qualitative Commentary From VRE Members / VRE-D To Confirm The Changes / Work Done On The VRE

The fact that the survey tool had been activated on the Alfresco system, and was operational, can be confirmed from comments sent via an e-mail from the B-VRE-M to the VRE-D on 8 December 2015: “B-R2 [name anonymised] and I am busy uploading onto Lime Survey via Alfresco. She has a number of questions. Would it be possible that we can see you some time for advice?” (e-mail, 8 December 2015, BD20).

Another e-mail sent by B-R2 [name anonymised] to the VRE-D on 4 February 2016 also confirms that the survey tool was operational via Alfresco: “Can I ask you to close the survey for me? Five people edited it for me, and there are a number of things I will need to change. I will do the changes and would like to ask that we Skype next week, just to do the final changes. I wish to finalise the link by the end of next week so that the parents can start completing it” (e-mail, 4 February 2016, BD21).

7.2.2.22 Implement Changes To The VRE

The changes mentioned in 7.2.2.21, i.e. activating the survey tool, was implemented on 4 November 2015, and the link to the questionnaire that was set up on the survey tool was finalized by 11 November 2016.

7.2.3 Summary Of The Formative Evaluation

The formative evaluation of Case Study A followed a process of PAR. Testing and prototyping of the VRE technological frameworks that were designed for this group, notes taken during meetings, training sessions with the members of these case studies, as well as e-mail correspondence between the VRE design team and the members of these VRE groups, were used as data collection methods. The first step was the identification of a case study (Case Study A). A Faculty that showed interest and concern about the RDM practices was approached with the possibility of hosting a pilot study, and a research group was identified that would be used as a pilot study. Next followed an exploration of the group’s processes, practices, and tools. After this, a meeting was held with the members of the group to determine their needs (in other words, a needs identification) with regards to the management of research data. Following this meeting, a document outlining the pilot project process was compiled, which suggested a bottom-up approach, as researchers are primarily responsible for the management of their own data. It was also decided that the pilot project would be hosted on a specific server on the Hatfield Campus of the University, with a proposed server of their own in the medium term, to ensure security of their data. At the same time a site was created for the group on a Moodle platform.

During a follow-up meeting with the members of the group, a demonstration was given by the researcher of this study and an IT specialist (later the VRE-D) of the files of a student researcher that had completed her studies, which had been uploaded on Moodle and DSpace. The student researchers were also showed how to structure folders in such a way that it would support the Research Lifecycle (see 3.4.2 and Figure 3.4). This was followed by a discussion on the needs of the group.

The outcome of the first formative evaluation led to the adaptation of Moodle as a VRE site, and the registration of members of Case Study A on the site. A link was also added to Google Drive for synchronizing purposes. A server was installed at the location where Case Study A is situated. These changes were followed by the first training session on Moodle for members of this group, on 19 September 2013.

During the second formative evaluation, it was decided to replace the Moodle instance of the VRE with an instance created on Alfresco, especially because it had a versioning function, which Moodle did not have. It was also seen as a more user-friendly system, with a number of positive functionalities as mentioned in 7.2.1.8. Case Study A's site containing their personal profiles, files and data were then migrated to Alfresco. These changes were followed by a second training session on 6 December 2013 to members of Case Study A, on the functionalities, procedures and processes in Alfresco.

A third formative evaluation with Case Study A dealt with a number of network, server, and external problems members had experienced, such as routing problems on the server, a problem with Google Docs where the code was changing all the time, duplication of folders, and network downtime that caused problems regarding access to the server. All these problems and issues had been addressed and resolved by the VRE-D by 15 September 2014.

The fourth formative evaluation focused on the establishment of a NAS storage device at the location where Case Study A was situated, the replication of the NAS device to one in the Library, the back-up of machines in the laboratory, the installation of a network point in the laboratory, upgrading of some of the machines in the laboratory, the installation of a 24-hour server room with four network points and two UPS's, the sorting out of a firewall problem, and the provision of access to a virtual machine environment to one of the student researchers, in order to test some of the software and processing power.

Case Study B was identified as an additional VRE group when the promotor of that group approached the researcher of this study and the VRE-D for the possibility of creating a VRE for her group. She mentioned that they were urgently looking for a system that could assist them in managing their research data in a more structured way. She had heard of the Alfresco instance that was created for Case Study A and was interested in a similar instance on the platform for her group of student researchers. The promotor then also attended the second training session that was presented to members of Case Study A.

The first formative evaluation for Case Study B flowed from a meeting held on 17 January 2014 to identify the group's needs with regards to a VRE. It was found that the group needed a workflow function between the promotor and the student researchers. They also needed versioning and archiving functions, as well as a survey tool. The VRE-D then created a site for the group on Alfresco, and registered members of the group onto the system. He also plugged a survey tool into the system. The meeting on 17 January 2014 was followed by a hands-on training session in Alfresco on 27 January 2014. The first instance for the group was implemented shortly after the training session, and members were encouraged to work through the system and play around with it.

An expansion and exploration of needs led to a second formative evaluation on 28 January 2014, of Case Study B's instance on Alfresco. During the second formative evaluation, a problem with the workflow of the system that was blocking the system from sending out e-mail notifications, a problem with logins to the system, and a problem with security settings that exposed data from members of Case Study A, were identified and corrected.

A third training session was arranged after a request from the B-VRE-M for two student researchers that were situated geographically far away from the campus. The one student was trained on 17 February 2014 and the other at the end of March 2014. The session included training on a survey tool that was plugged into the system.

The training session was followed by a third formative evaluation on 19 January 2015, where the VRE-D helped the B-VRE-M to download an Alfresco application (app) onto her laptop, and demonstrated how to take photos and videos and record audio through the Alfresco app, by using the device's microphone and camera. He also added a calendar dashlet to the group's site.

A fourth formative evaluation followed next, where the VRE-D investigated a problem that arose with access to the system and notifications via e-mails. He executed an upgrade of the system hoping that this would solve the problem, but the problem persisted. The VRE-D then investigated the problem further, and found that there was a problem with the notification function itself on the Alfresco system. He subsequently fixed this.

A fifth formative evaluation flowed from e-mails sent on 19 May 2015 and 25 May 2015 from the B-VRE-M to the VRE-D. The workflow function in Alfresco was then demonstrated to the head of her academic department, and two co-study leaders from other institutions were registered on the Alfresco system with rights to monitor specific student researcher's progress.

A sixth formative evaluation identified the need for a survey tool, which resulted in adaptation of Alfresco by plugging in a survey tool (created with LimeSurvey) that could be used by student researchers in Case Study B.

7.2.4 Schematic Figures Of Development Of VRE's For Case Study A And Case Study B

The PAR process through which VREs for Case Study A and Case Study B were developed, is illustrated by means of self-reflecting spirals (see 6.2.2.3) called Figures 7.4 and 7.5. As mentioned in 6.2.2.3, these cycles can be illustrated as cycles that follow each other on a timeline (see Figure 6.3), but in reality, it looks more like a spiral with concentric flows (see Figure 7.3).

In each case study, the PAR process consisted of a number of concentric cycles. The members (see blue line in Case Study A and red line in Case Study B) identified their own needs, received training, used the VRE platforms and tested it. The role of the VRE design team is indicated by means of a yellow line, and consisted mostly of identification of the needs of the VRE group members, the training of members of these groups, adaptations of these VRE platforms, and implementation of changes to these VRE platforms. The identification of a VRE for Case Study B originated from the second training session in Case Study A. This training session also then formed the second training session of Case Study B.

Figure 7.4: Formative Evaluation – Case Study A

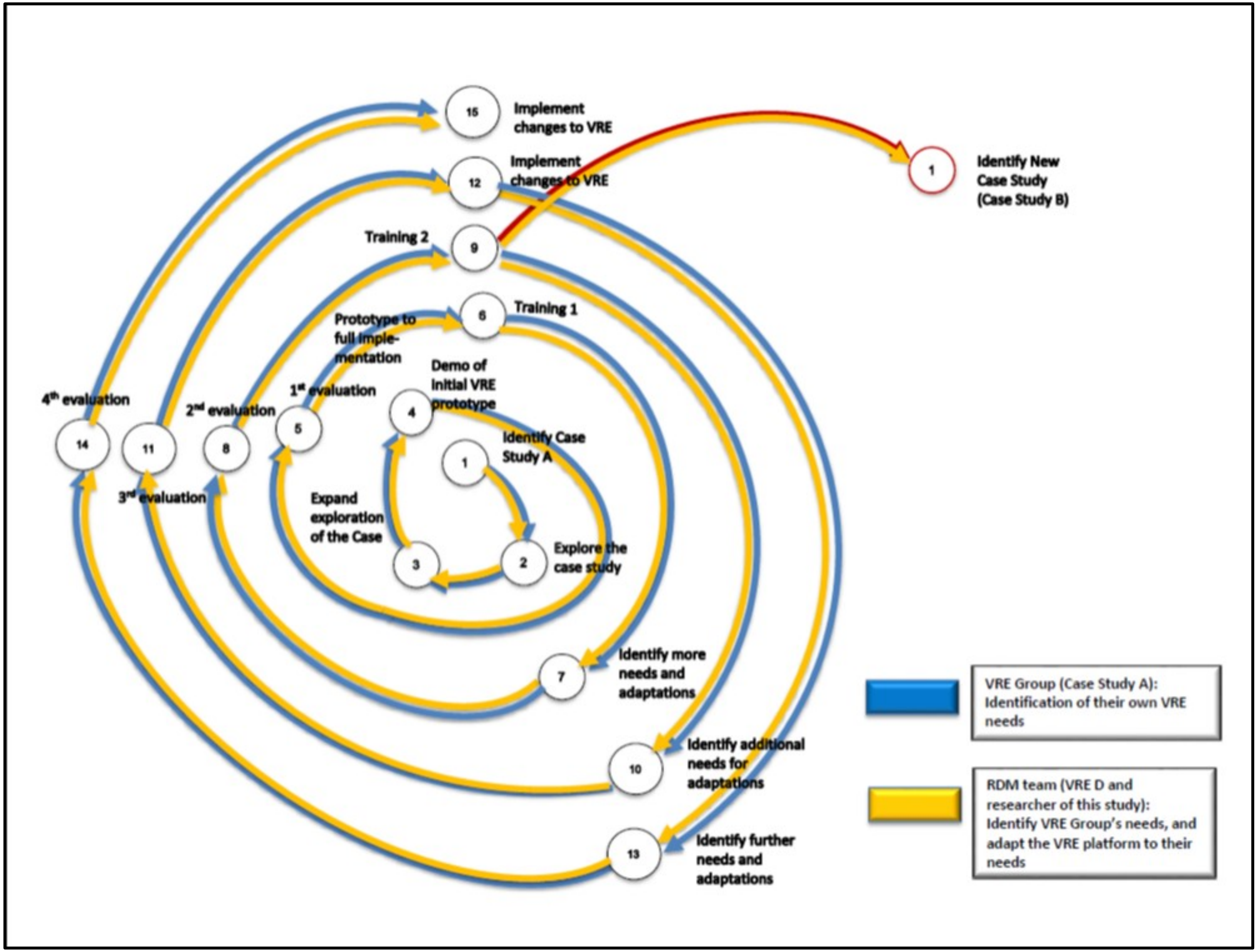
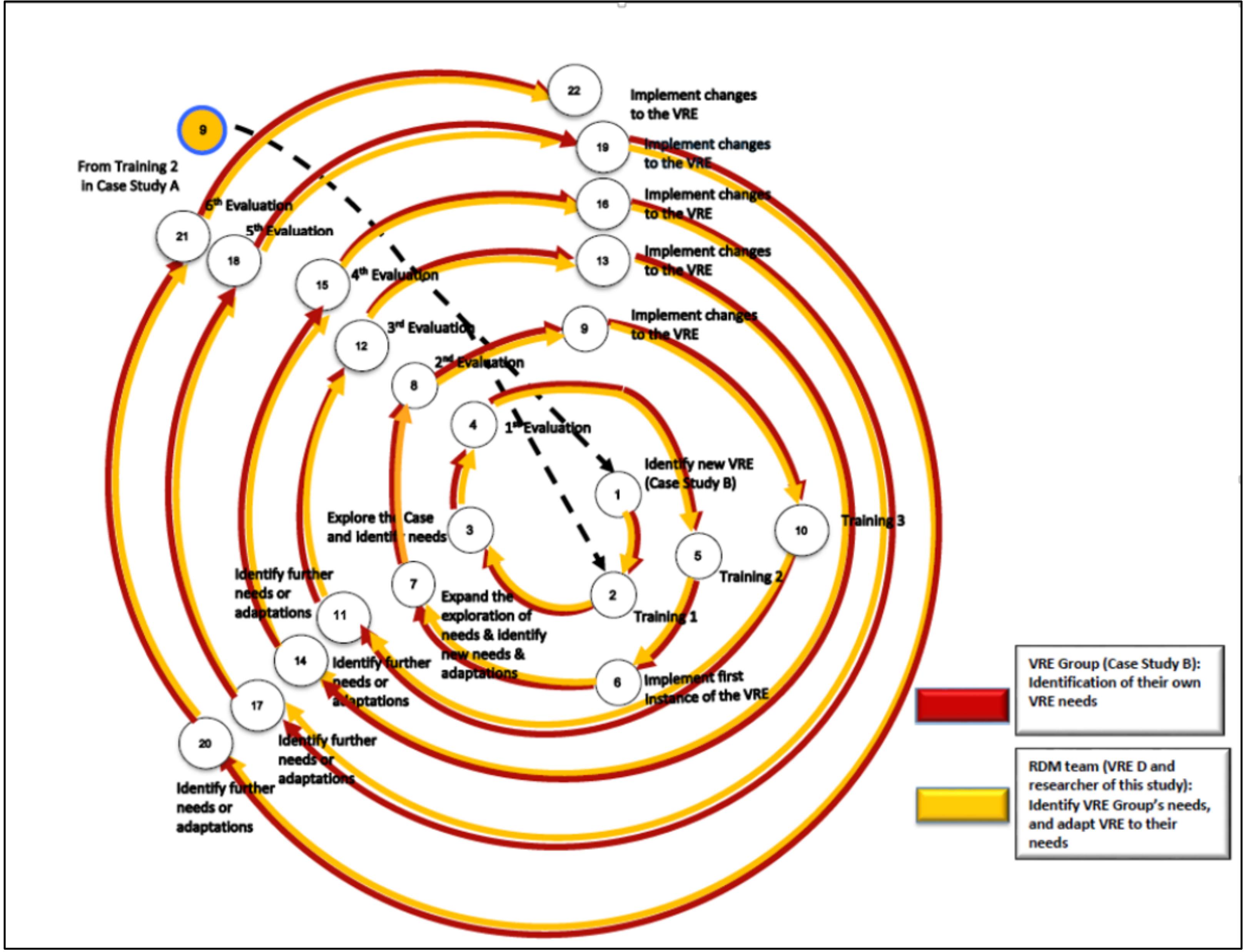

15
Implement changes to VRE
1
Identify New Case Study (Case Study B)
12
Implement changes to VRE
Training 2
9
6
Training 1
Prototype to full imple-mentation
Demo of initial VRE prototype
1st evaluation
4th evaluation
2nd evaluation
4
14
11
8
5
Identify Case Study A
1
3rd evaluation
Expand exploration of the Case
3
2
Explore the case study
7
Identify more needs and adaptations
10
Identify additional needs for adaptations
13
Identify further needs and adaptations
VRE Group (Case Study A): Identification of their own VRE needs
RDM team (VRE D and researcher of this study): Identify VRE Group's needs, and adapt the VRE platform to their needs

Figure 7.5: Formative Evaluation – Case Study B



22
Implement changes to the VRE
9
From Training 2 in Case Study A
19
Implement changes to the VRE
6th Evaluation
5th Evaluation
16
Implement changes to the VRE
21
18
15
4th Evaluation
13
Implement changes to the VRE
12
3rd Evaluation
9
Implement changes to the VRE
2nd Evaluation
8
4
1st Evaluation
Identify new VRE (Case Study B)
1
Explore the Case and Identify needs
3
10
Training 3
5
Training 2
Identify further needs or adaptations
7
2
Training 1
11
Expand the exploration of needs & identify new needs & adaptations
14
Identify further needs or adaptations
6
Implement first instance of the VRE
17
20
Identify further needs or adaptations
Identify further needs or adaptations
🟥 VRE Group (Case Study B): Identification of their own VRE needs
🟨 RDM team (VRE D and researcher of this study): Identify VRE Group's needs, and adapt VRE to their needs

7.2.5 Conclusion About The Formative Evaluation

The formative evaluation of Case Studies A and B led to the identification of functionalities and components that were seen as important and were well-used by members of these groups; for example, archiving and back-up of data, versioning, the workflow function, e-mail notifications of actions happening on the VRE, and synchronization of files on a desktop computer via an app to the VRE. A number of new functionalities were also identified, such as the adding of a link from the VRE to Google Drive, a survey tool, an additional storage device for Case Study A, the replication of this storage device to one in the Library, the backing-up of machines in the laboratory, the connection of these to the network, and the provision of access to a virtual machine environment to one of the student researchers to test some of the software and processing power.

The formative evaluation also revealed that the Moodle instance of the VRE was not meeting all the needs of the researchers and led to the migration to an instance of a VRE created on Alfresco. The formative evaluation eventually led to the recommendation by the VRE-D in the summative evaluation, for a future migration of the VRE to HUBzero (a software platform specifically created for VREs).

7.3 SUMMATIVE EVALUATION

The summative evaluation, as mentioned in 6.2.2.5 (b), consisted of semi-structured interviews as the instrument, because the “face-to-face interaction between the interviewer and an interviewee” offered an “understanding of experiences” or circumstances as described by the interviewee in his or her own words (Schurink, 1998: 20). The discussion points of the semi-structured interviews covered the following aspects:

- What is RDM?
- What is a VRE?
- What is the current state of VRE research in the world?
- What are the generic components / tools that make up a VRE?
- How does a VRE support a research cycle?
- Why should RDM be an essential component within a VRE?

- To what extent can the components identified within Chapter 1, Section 1.2.4 be formalised into a conceptual framework and where would RDM as component be placed?
- To what extent can this model be generalised for use in other environments?
- To what extent can guidelines be developed for such a conceptual VRE model?

7.3.1 Questions Of The Interview Schedule

7.3.1.1 Questions To Postgraduate Student Researchers Participating In The VRE

Section A: Questions On The Virtual Research Environment (VRE)

(1) This VRE project exists 18 months, when did you join this project? (number of months)

Each of the respondents was assigned a code (given in brackets) for easier analysis later in this chapter. For example, when referring to Researcher from Case Study A, the code would be A-R1.

Case Study A	From the Start (May 2013)	Later
Researcher 1 (A-R1)	✓	
Researcher 2 (A-R2)	✓	
Researcher 3 (A-R3)		✓ (March 2014)
Researcher 4 (A-R4)	✓	
Researcher 5 (A-R5)	✓	
VRE Champion (A-VRE-C)	✓	
VRE Manager (A-VRE-M)	✓	
VRE Designer (VRE-D)	✓	

Case Study B	From the Start (beginning January 2014)	Later
Researcher 1 (B-R1)	✓	
Researcher 2 (B-R2)		✓ (March 2014)
Researcher 3 (B-R3)	✓	
Researcher 4 (B-R4)	✓	
VRE Manager (B-VRE-M)	✓	
Librarian (B-L)		✓ (End of January 2014)
VRE Designer (VRE-D)	✓	

The answers to this question showed that one of the researchers (A-R3) in Case Study A had only joined the VRE at a later stage, in March 2014, which could potentially have had an influence on the answers given in the rest of the interview, as this respondent did not have as much time exploring and using the system as the other respondents, and may not be as familiar with all the tools of the VRE. In Case Study B, all the answers from respondents revealed that all had been involved in the VRE from its inception early in January 2014, except one of the postgraduate researchers (B-R2), who joined the site only in March 2014, and the librarian (B-L) who joined the site only at the end of January 2014. As in the case of Case Study A, one could expect that this would have had an influence on the answers given to questions asked during the interview.

The VRE designer and the researcher of this study were involved in both these groups from the start.

(2) What is your role within the VRE project?

The answers received in Case Study A provided a good overview of the different roles that had been assigned to each of these respondents, and showed that each had a clear idea of the role that they were expected to play in the VRE. The five postgraduate researchers described their roles as being either student researchers, or postgraduate researchers, or senior scientists, or just users of the VRE. Two stressed that they had used the VRE to manage or upload their data or information. The VRE Manager (A-VRE-M), who was also the promotor of these students, had a very limited view of his role. He described his role as seeing to it that the researchers' data were stored in a secure place, in this case the VRE, for 14-15 years. He was also responsible for ensuring that protocols sent through to the Faculty's Ethics Committee were accompanied by a declaration that stipulated where the data would be kept, and for how long. He did not, however, mention that he monitored the uploading of the student researchers' data, or verification of the quality of this data constantly. The VRE Champion (A-VRE-C) saw her role as being the primary contact in the group as well as its administrator. The VRE designer (VRE-D) was seen as the primary administrator of the VRE.

The answers in Case Study B showed that each of the members had a clear idea of what their roles entailed. The VRE Manager (B-VRE-M) indicated that she had been the supervisor of the student researchers. The four student researchers described themselves as students that had been using the VRE to manage (store, share and access) their data. Further discussions though revealed that they used the VRE for other functionalities as well, such as communication (see question 5), sharing data (see question 5), sharing files (see question 8), running workflows (see question 8), versioning (see question 8), placing their files under categories (see question 8), keeping track of site activities (see question 8), and conducting surveys (see question 8). The Librarian (B-L) described herself as an information specialist that had been contributing to the VRE by doing information searches for students / researchers. She had been uploading the results of these searches as well as specific articles that had been requested. She had been using it as a shared environment with more than one postgraduate student.

(3) Does the VRE project you belong to, focus on one discipline, or would you say it is multi-disciplinary?

In 5.3, it was mentioned that VREs have an interdisciplinary nature that allows for the gathering of data and approaches from different disciplines to create new research findings (Carusi and Reimer, 2010: 23, Fraser 2005). The two case studies as mentioned in 6.2.2 focused on two different disciplinary areas. Case Study A had been using natural science-oriented data and laboratory / experimental methods, whereas Case Study B had been using human-oriented data and survey instruments as data collection method.

The answers to this question showed that even within each of the case studies, the respondents had differing views. The majority of the respondents in Case Study A were of the opinion that the VRE they were part of, catered for more than one discipline. One researcher (A-R4) saw it as one discipline, but cutting across other disciplines or fields, while another researcher (A-R5) saw it as being just one discipline. The A-VRE-C saw the case study as mostly “science-based”, but very diverse.

In Case Study B, which dealt with human-oriented data, the majority of respondents saw their VRE as catering for one discipline, whereas one researcher (B-R2) saw her own project as being multi-disciplinary. Another researcher (B-R3) expressed the opinion that the other members of the VRE had been focusing on different subject matters within a specific field, unlike her project, where she had to collaborate with another discipline. This, according to her, would perhaps make her project the first multi-disciplinary project in the VRE. This showed that there had been a need in Case Study B for a multi-disciplinary VRE, even though minimal.

The requirement for the provision of multi-disciplinarity in the VREs could have had an impact on the manner in which these VREs and their components developed, as could be seen in the answers to the rest of the questions.

(4) Is the focus of the project topic-centred or technology-driven? Please explain.

Case Study A		
Respondent	**Technology-Driven**	**Topic-Centred**
A-R1	✓	
A-R2		✓
A-R3	✓	
A-R4	✓	
A-R5		✓
A-VRE-C		✓
A-VRE-M		✓

Case Study B		
Respondent	**Technology-Driven**	**Topic-Centred**
B-R1		✓
B-R2		✓
B-R3		✓
B-R4		✓
B-VRE-M		✓
B-L		✓

In Case Study A, four of the respondents (A-R2, A-R5, A-VRE-C, A-VRE-M) saw the VRE as being driven by the topics of the research, while three (A-R1, A-R3, A-R4) were of the opinion that the VRE was driven by the technology itself. This meant that four researchers were of the opinion that the technologies in the VRE served only as tools that could support or enhance their research and that the technologies did not drive their research. In other words, the VRE had been designed around their needs. Three of the researchers felt that the technology of the VRE had been driving the way they did their research, in other words, they had to adapt the way they did their research to the way the various components and tools in the VRE worked.

In Case Study B, all the respondents expressed the view that their VRE had been driven by the topics of the research itself. This meant that they saw the technologies in the VRE merely as tools that could support or enhance their research, in other words, the VRE and its components had been designed around the needs in the various projects.

(5) What do you understand under the term "Virtual Research Environment" (VRE)?

Specific themes could be identified from the respondents' answers to this question. These have been listed in column 1 of Table 7.1 and have been matched with findings from the literature in column 2, where only a cross reference to the section where the issue is discussed, is given, and with the direct quotations from the respondents in column 3. This is followed by a discussion of the information found in the table.

Table 7.1: Themes that describe the concept Virtual Research Environment

Theme	Literature	Quotation
Online / Digital system / framework	A VRE consists of a common, flexible, technological and collaborative framework (see 2.2.8.1).	A-R1: "The concept means to me that the research is all online and easily accessible and made accessible to the environment." A-R2: "I would describe it as an electronic space." A-R3: "A computer-based system that contains all information on the research project." A-R5: "A research environment that is not really physical, but within an IT domain"

		B-R3: “If I think of virtual it means something online. So it’s something that a person is connected to online and are able to share the information on one central platform.” B-L: “An online environment, that is distance-based.”
Cloud-based	The hardware component of a VRE consists of four components, of which one is cyberinfrastructure, which includes local networks (e.g. servers), the national backbone, and international infrastructure (e.g. *cloud services*) as tools to assist in accommodating the vast amounts of data that will need to be managed (see 3.5.7.2).	B-R1: “I understand it as a type of *cloud* or database, in which we can share articles and information with each other on an aspect / topic that interest all in the group.” B-R2: “I see this as a *cloud*, similar to an iCloud. A cloud that you can access wherever you are, and on which you can do your research directly.” B-R5: “A VRE can be used to communicate using information technology, to save data, or to share data in a *cloud* or similar.”
Storage / Archiving	A VRE can provide an easy to use technological framework where researchers can secure the short-term storage of their data, and by integrating an architecture for data management within a VRE, the matter of preservation of research data can be addressed (see 5.3).	A-R1: “*Store* it in terms of almost like a library where you have a space where you can *store* everything and access it when you need it.” A-R2: “An electronic space in which data gets *stored* and updated.” A-VRE-C: “Where information is always *stored* and backed-up and where it is kept safely.” B-R1: “I also see the VRE as a place where I can *store* my data for safekeeping.” B-R2: “I can also *save* the data that I collect onto that.” B-R5: “A VRE can be used to communicate using information technology, to *save data*, or to share data in a cloud or similar.”
The possibility to add plugins	A VRE can be described as a flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts, etc.) interoperating with each other, can be plugged (see 2.2.8.1).	A-R2: “If you want, you can add a more interactive component.”
Access to information and data	“A VRE should provide an effective personalised access point to information, experts, knowledge, collaboration tools and computational resources” (Van Deventer et al., 2009) (see 2.2.8.3).	A-R1: “Space where you can store everything and access it when you need.” A-R4: “I see it in in terms of data that I can access anywhere in the world from a reliable sustainable source.”

Sharing of information and data	VREs can be used for analysis and processing of data, annotating data collaboratively, and sharing of data with peers (see 2.2.8.1).	A-R1: “My group for instance we have been able to *share* all these documents.” A-R5: “A VRE can be used to communicate using information technology, to save data, or to *share* data in a cloud or similar.” B-R1: “I understand it as a type of cloud or database, in which we can *share* articles and information with each other on an aspect / topic that interest all in the group.”
Collaboration and interaction	“A key characteristic of a VRE is that it facilitates collaboration amongst researchers and research teams” (Brown, 2012) (see 2.2.8.3). Another characteristic of a VRE is that “a VRE system should be able to act as communication platform” (Yang and Allan, 2006a: 453; Wilson, et al., 2007: 290) (see 2.2.8.3).	A-R2: “if you want, you can add a more interactive component.” A-R4: “I see it as a source that should allow me to *interact* with people no matter where they are, through a central point.” A-R5: “A VRE can be used to *communicate* using information technology.” B-R4: “The VRE functions as a *supporting network* for all that are involved.” VRE-D: “So it would be a one-stop solution, where they sign on and all their data is there, all their tools are there, *collaboration aspects*, etc.”
VREs can stretch across organisational and geographical boundaries	A VRE can support research performed by researchers in multidisciplinary contexts and across organisational and geographical boundaries (see definition in 2.2.8.1). VREs can also bring researchers that are geographically dispersed together, by providing the necessary tools that will enable them to work more intensively on a project than would have been possible in a once-a-year meeting (see 5.3). This aspect of bringing together geographically dispersed researchers is also mentioned by Anderson, Dunn and Hughes (2005: 516) (see 5.3).	A-R4: “I see it in terms of data that I can access *anywhere in the world* from a reliable sustainable source. I also see it as a source that should allow me to interact with people *no matter where they are*, through a central point”. A-VRE-M: “I see it as an ecosystem, which has multiple components. A local component at the Institute, a bigger institutional component within the University, and a national component.” B-L: The Librarian described it as “an online environment that is *distance-based*, and allows for the sharing of information.”
Integrated with researchers’ everyday activities	Most VREs are also fully integrated with the research process (cycle) (see 5.3)	A-VRE-M: This respondent sees VREs as part of what researchers do every day – “it’s just like breathing every day.” A-VRE-C: A VRE is “almost like a virtual lab book.” It is “*something that you use every day*, that you update constantly, and where information is always stored and ‘backed-up’, and

		where it is kept safely.” She also sees it as something that should ideally be *integrated in the laboratory.*
Management of data	See discussion on RDM as an essential component in 5.3 of this study.	A-VRE-M: This respondent sees the management of data as one of the major drivers of a VRE.
Multiple components	The idea that a VRE can have multiple components can be seen in 5.4 of this study.	A-VRE-M: This respondent sees it as “an ecosystem, which has multiple components.” According to him, this ecosystem consists of a local component at the Institute, a bigger institutional component within the University, and a national component that could tap into the CHPC and DIRISA, and an international component.

The following paragraphs further expand on the issues raised in Table 7.1.

In Case Study A, A-R1 saw a VRE as an online space where one could do one’s research, while A-R2 described it as an electronic space, A-R3 saw it as a computer-based system that contained all information on the research project, and A-R5 saw it as a research environment that is within an IT domain. Respondents in Case Study B answered in a similar vein. B-R3 mentioned that the concept of ‘virtual’ implied that it meant something online, while the B-L described it as an online environment that is distance-based. This corresponded to the researcher of this study’s idea of a technological collaborative framework to support and enhance large and small-scale processes of research, which are performed by researchers in multidisciplinary contexts and across organisational and geographical boundaries, as mentioned in 2.2.8.1.

The aspect of ‘sharing’ also corresponds to the idea of a ‘collaborative’ framework, mentioned in 2.2.8.1. A-R1 further highlighted the ideas of ‘storage’ of and ‘access’ to documents and other electronic objects, e.g. data. Although these ideas had not been mentioned directly in the definition in 2.2.8.1, they could be seen as part of the processes of research performed by researchers in a VRE. The description of it being ‘almost like a library’ had been very limiting, however. This had been pointed out by Wusteman (2009: 170), referenced by the researcher of this study in 2.2.8.1. According to Wusteman (2009: 170), a VRE is more than a digital library, or portal to a range of digital activities. The description of a VRE by A-R3 as a computer-based system that contained all information on the project sounded the same as this limited idea of it being just a digital library. A-R2 described a VRE as a space where one could upload and update one's data, and also mentioned one of the characteristics of a VRE, of adding (or plugging in) other interactive components. This was in line with the researcher of this study's definition of a VRE in 2.2.8.1, where a VRE is described as a flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts, etc.) interoperating with each other, can be 'plugged'.

A-R4 added the idea of interacting with others by using a VRE, and A-R5 added the idea of communicating and using information technology, to the idea of saving and sharing of data. This corresponds with the idea of collaboration between researchers as mentioned in this researcher's definition in 2.2.8.1. The idea of a VRE being a communication platform is also mentioned by Yang and Allan (2006a: 237) and Wilson, et al. (2007: 290) as being a characteristic of a VRE (see 2.2.8.3). The sharing aspect of a VRE was also highlighted by A-R1, when she mentioned the sharing of documents. B-R1 in Case Study B added the idea of sharing of information and articles. The sharing of data was shown in 5.3 to be supported by Filetti and Gnauck (2011: 237) as a key element of a VRE. This sharing aspect is also mentioned by Carusi and Reimer (2010: 19), as referenced in 2.2.8.1.

The idea of using a VRE to store, archive or save data was mentioned by respondents A-R1, A-R2, A-R5, as well as the VRE-C. This was also echoed in Case Study B by three researchers (B-R1, B-R2, B-R4), the B-VRE-M, and the B-L, and is in line with Brown (2013) and Robertson Library (n.d.) that identified the use of a VRE to save data, or secure data collaboratively, as a key characteristic (see 5.3). This could also be seen in Carusi and Reimer's (2010: 19) Virtual Research Environment Collaborative Landscape Study, which showed that integrating an architecture for data management within a VRE can address the issue of preservation of research data, by providing a ready-to-use platform where researchers can secure the short-term 'storage' of their data (see 5.3).

The idea of a VRE running specifically on a cloud system, was mentioned as part of the hardware components layer proposed by the researcher of this study in 3.5.7.2, and is listed specifically under the cyberinfrastructure component within this layer.

The opinion that a VRE can stretch across organisational and/or geographical boundaries was brought to the surface by A-R4, when he mentioned that he sees a VRE "in terms of data that I can access anywhere in the world from a reliable sustainable source" and also "as a source that should allow" him "to interact with people no matter where they are, through a central point." The B-L described it as an online environment that is distance-based and allows for the sharing of information. This is supported by the researcher of this study's definition of a VRE in 2.2.8.1, where he indicates that a VRE "can support research performed by researchers in multidisciplinary contexts and across organisational and geographical boundaries." This was confirmed by Carusi and Reimer's (2010: 22), referenced in 5.3, who state that VREs can bring researchers that are geographically dispersed together by providing the necessary tools that would enable them to work more intensively on a project than would have been possible in a once-a-year meeting. This aspect of bringing together of geographically dispersed researchers was also mentioned by Anderson, Dunn and Hughes (2005: 516) (see 5.3).

The aspect of access to information and data through a VRE was brought to the fore by A-R1, when she mentioned that a VRE is a space where one can store everything and access it when one needs it. It was also raised by A-R4, who sees a VRE in terms of data that he can access anywhere in the world from a reliable sustainable source. This was in line with Van Deventer et al. (2009), who are of the opinion that a VRE should provide an effective personalised access point to information, experts, knowledge, collaboration tools and computational resources (Van Deventer et al., 2009) (see 2.2.8.3).

Another theme that came to the fore through the researchers' answers was the idea that a VRE should be integrated with researchers' everyday activities. The A-VRE-M saw VREs as "part of what researchers do every day" and the A-VRE-C saw a VRE as something that one uses every day, that one updates constantly and something that should ideally be integrated in their laboratory.

The A-VRE-M proposed the idea that a VRE has multiple components. He saw it as an ecosystem, which has multiple components, namely a local component at their department, a bigger institutional component within the University, and a national component that could tap into the CHPC (Centre for High Performance Computing), situated in Cape Town, and DIRISA (Data Intensive Research Initiative of South Africa), situated in Pretoria, and an international component. This corresponded with a proposal made in 5.4 by the researcher of this study, of a possible VRE with multiple components. Many of these components mentioned by the A-VRE-M are implied, but packaged differently.

The A-VRE-M also saw the management of data as one of the major drivers of a VRE. In other words, it is a very important component of a VRE, which is in line with the review of literature in 5.3, which showed that RDM is indeed an important component within a VRE.

(6) Were you afforded an opportunity to give input into the design of the VRE? If so, what type of input did you give? (If you joined the VRE project after it was launched, were you afforded an opportunity to comment on the current design?)

The aim of this question as indicated in 6.3 was to establish if these VRE projects (case studies) were implemented in a top-down approach or if the respondents were given an opportunity to provide input into how the VREs should look like.

The majority of respondents from Case Study A remembered just one meeting that was held with them, where they could give input. A-R2 remembered that their group had a couple of meetings with the VRE-D, where various flavours of VREs were presented to them, but she had not been sure how much of that was factored into the system that they were using at the time of the interview. She could not remember if they had given input, or what the VRE-D had done in terms of the coding, to customise it for them. She further expressed the need for future development of the VRE to link it with EndNote or RefWorks. A-R3 joined the VRE at a later stage, which meant that she could not give input on the start of the VRE. She had been under the impression that it was a set system, and had not been aware that she could provide input, even at a later stage. A-R4 remembered that a meeting had been held where the broad options were discussed and where they as a group indicated what they wanted to achieve. He specifically remembered that the group indicated that such a system would be great for managing their data. The VRE-D then designed something around those needs. A-R5 felt that the members of the group had been allowed to give input even though it was minimal. According to her, it had been a continuous process. They were given the freedom to share their thoughts on what would be the best way, and this included sharing their challenges and providing suggestions. The main input though, according to her, had been provided by the A-VRE-C. The A-VRE-M mentioned that they had provided input in the beginning, but that the process of giving input had been a continuous process. The A-VRE-C indicated that she had been included from the beginning in the design process. She had given input on how it would be implemented and how it would be structured in the best way within the Alfresco framework.

Case Study B was started after the research promotor (B-VRE-M) of the specific research group heard of the potential for usage of the VRE by Case Study A (see 7.2.2.1). Alfresco was deemed by the B-VRE-M as a good system for the group to use as a VRE tool. In other words, Alfresco had already been the system of choice, when the group started. This led to some of the respondents (B-R1, B-R2, B-L) feeling that they had not provided input into the design of the VRE, and that the system was introduced in a top-down fashion. If one however scrutinizes the rest of the answers given by B-R2, it would seem that they did have a session where the VRE-D explained the system very clearly and where they could ask questions and give opinions on it. This indicates that they did have an opportunity to provide input into the functionalities they wanted to use, although not in the choice of the system that they would want to use. The B-L indicated that she had not really been involved in giving input, but that she did give input in helping one of the researchers in structuring the content of her information into folders, etc.

(7) Did you receive training to use this VRE? If so, what type of training did you receive?

In Case Study A, the majority of respondents indicated that they had attended at least one formal training session presented by the VRE-D. A-R3 joined the VRE at a later stage; consequently, she did not receive the same formal training as the rest of the group. She was shown the basics of the system by the A-VRE-C, but learned the rest by trial-and-error. A-R5 could also not attend the formal training session. The fact that these two respondents could not attend had an impact on the way they experienced the VRE, as demonstrated in answers to later questions in their respective interviews.

A-R1 indicated that there had been a formal hands-on-training session at the Hatfield campus, followed by a Microsoft PowerPoint session at the campus where she does her research. This had been followed up by support and assistance from the A-VRE-C from time to time, as needed. A-R2 declared that they had a number of informal training sessions with the A-VRE-C, but also a formal session with the VRE-D. During the session, the VRE-D had shown them how to use the VRE. This included, for example, a demonstration on how to upload, how synchronising and versioning works, and how to set up workflows, etc. A-R4 mentioned that they only had one formal training session presented by the VRE-D on the Hatfield campus. According to him, they had no follow-up training. The A-VRE-C, according to him, had played a key role in kick-starting the VRE and advancing the VRE. She had also provided the necessary advice to other members of the VRE. The A-VRE-C mentioned that they did have the initial training session with the VRE-D, but that this had been followed up by continuous training sessions between her and the VRE-D. The A-VRE-M had only attended one formal training session at the start of the VRE.

In Case Study B, the B-L indicated that she had not really received training, as the session that she had attended, had only been a Microsoft PowerPoint presentation by the VRE-D. She had not been able to attend the hands-on-session (see 7.2.2.5) that had also been conducted by the VRE-D, because she had only joined the VRE after this session had taken place. This had an influence on the way she perceived the system and it influenced her answers to other questions later in the interview. B-R1 indicated that they had only had one session of training (see 7.2.2.5), and further indicated that this training had been very minimal. She did however stress that she thought the system could do much more, but had not had enough time to explore these other functionalities by herself. B-R2 mentioned that they had received training, but did not indicate when or how many sessions. She added that when they had questions, the VRE-D had always been available via e-mail to help. B-R3 stated that they had only one training session (see 7.2.2.5) in 2014, where they were given an orientation to the site, but that she had not used it yet, which meant that she could not gauge whether the training had been sufficient. The response by B-R4 was that they had attended a training session (see

7.2.2.5), which had been in the form of a simulation of what they would do in a real-life situation. She also stressed that it had been a comprehensive training session. Her perception differed from that of B-R1, and could perhaps indicate a difference in their levels of computer literacy or readiness. The B-VRE-M confirmed that they had received training on the VRE system as a group, but that she had spent about two training sessions individually with the VRE-D in addition to that, to improve her understanding and know-how of the system.

The responses received to this question showed that training was needed, especially in cases where respondents were not very computer literate. This meant that the system was not as user-friendly as one would have wanted it to be.

(8) Which of the current tools/components in the VRE are you currently using? Please indicate why you are you using each of these tools/components, and not the others?

The purpose of this question was to determine which components had been used, which could potentially be included in a conceptual framework model.

- **Create A Site**

A site was created on Alfresco for all the respondents of Case Study A and another site was created for all the respondents of Case Study B.

- **Edit Your Profile**

In Case Study A, five of the respondents (A-R1, A-R2, A-R3, A-R4 and A- VRE C) mentioned that they had edited their profiles, while two of the respondents (A-R5 and A-VRE-M) mentioned that they had not edited their profiles. A-R5 indicated that she had not been using it because it takes time, while the A-VRE-M mentioned that he had not been using the system extensively.

In Case Study B, three of the respondents (B-R2, B-R4, B-L) indicated that they had edited their profiles, while three (B-R1, B-R3, B-VRE-M) indicated that they had not. B-

R3 stated that she had not edited her profile yet. The reason for this, according to her, was that she had been waiting for audiology equipment from Denmark, and therefore had not been using the system yet. She had only gone into the system to see where files could be uploaded. The B-VRE-M mentioned that she had not been prompted by their situation to use it. Her answer could be indicative of the size of the membership of the VRE, which was quite small. It might not have been essential for them to edit their profiles, as the members of this group had known each other.

- **Search**

Two of the respondents (A-R1, A-VRE-C) in Case Study A indicated that they had used the search component, while five respondents (A-R2, A-R3, A-R4, A-R5, A-VRE-M) mentioned that they had not. A-R1 indicated that she had used the search component a couple of times in the beginning, but now that she knew where things were, she had not been using it as often. She had used it to search for files and for people. The A-VRE-C mentioned that she had used the search component under 'documents' to search, for example, for specific key words of a laboratory sample that they had backed up.

In Case Study B, two of the respondents (B-R1, B-R4) said that they had used the search component and two respondents (B-R2, B-R3) indicated that they had not used it. The B-VRE-M mentioned that the reason she had not used the search component was because the VRE project had been very small, and that she knew everyone in the group and also knew what the members were doing. This could probably also have been the reason why so many of the other respondents in Case Study A and B had not been using the search component.

- **Site Calendar**

In Case Study A, only one respondent (A-R2) revealed that she had used the calendar briefly in the beginning, whereas six of the respondents (A-R1, A-R3, A-R4, A-R5, A-VRE-C, A-VRE-M) divulged that they had not made use of the calendar. The reason A-R2 had only used the calendar briefly, and then stopped using it, was because no-one else had been using it. The calendar component is a tool that can only be successful if everyone is using it. It does not, for example, make sense to schedule an event on it, when no-one else sees it. Some of the other respondents (A-R1, A-R4 and A-R5) complained that the calendar component was not part of their daily practices or routines, and that it could only be used once one had logged into Alfresco. A-R1 and A-R4 indicated that there were other tools that they had been using as part of their daily practices, for example Google Calendar. These, however, were not linked to the Alfresco system and did not synchronise to the system. The A-VRE-C mentioned that the social aspect of the VRE was not functional yet.

In Case Study B, the answers of respondents revealed that none of them had been using the calendar function. The B-L stated that she had not been aware of the calendar, which is again the result of her not receiving all the training that the rest of the respondents had received. The B-VRE-M communicated that she had wanted to use the calendar in the past, and might still use it in future (see also in 7.2.2.12).

The main reasons for non-use of the calendar in Case A and B in summary then, seems to be non-awareness of the tool, the fact that it had not been integrated with their work processes, and the availability of other tools outside the VRE, which they had been more familiar with.

- **My Discussions**

The interviews revealed that none of the respondents in Case Study A used this component. The reason for this was articulated by A-VRE-C, when she mentioned that the social aspect of the VRE had not been developed yet. This corresponds with what was discussed in 7.2.1.5, when the initial VRE prototype was demonstrated. The group at that stage indicated that they did not need a lot of additional features or trimmings such as social media features. In Case Study B, three of the respondents (B-R2, B-R4, B-L) mentioned that they had been using it a few times, but three respondents (B-R1, B-R3 and B-VRE-M) indicated that they had not. B-R4 disclosed that she had been using it but not to its full potential and the B-L stated that she had used it once or twice. These answers could signify that the social aspect of the VRE in Case Study B had been developing faster than in Case Study A.

- **Following**

One respondent (A-VRE-C) in Case Study A stated that she had been using this component, but six of the respondents (A-R1, A-R2, A-R3, A-R4, A-R5 and A-VRE-M) revealed that they had not been using it. A-R3 indicated that she had not been aware of the component, which is probably due to the fact that she had only joined the group at a later stage and had not received the training that the rest of the group had undergone. A-R5 divulged that she followed activities on the VRE, but not a specific person on the VRE through the 'Following' component. The reason for non-use could again be related to the social aspect of the VRE that had not been developed by the group as yet, as mentioned by the A-VRE-C under 'My Discussions'. The A-VRE-C was revealed to be the only respondent that had been following other people, and this should add value to the role that she had been playing in the VRE as VRE Champion. The 'Following' component could enable her to stay informed about actions taken by each of the other respondents in the VRE, and to monitor activities.

In Case Study B, the pattern looked very similar to that of Case Study A. One of the respondents (B-VRE-M) indicated that she had been using the 'Following' component, while five (B-R1, B-R2, B-R3, B-R4, B-L) mentioned that they had not been using it. B-R2 revealed that she had not been using this component, because she was not ready yet for the social aspect of the VRE. She mentioned that she had not even been using social media such as Facebook. B-R4 said that they as members of the VRE were already linked to each other on the site and did not see the need to follow each other. The A-VRE-M, however, stated that she had been using the 'Following' component to connect with one of her very active students. This had enabled her to stay informed about everything that this student had been doing.

In Case Study A, the A-VRE-C also at times fulfilled the role of VRE Manager on behalf of the A-VRE-M, while in Case Study B, the B-VRE-M also fulfilled the role of VRE Champion. In order to keep abreast of what was happening in their VREs, the 'Following' component could have provided these two respondents with a valuable monitoring tool. It is therefore not surprising that in both these case studies, it had been these two respondents who had been using it.

- **My Files (Drag & Drop, Upload Files, Create Folders)**

This component was shown to be very popular with the respondents and all the respondents of both case studies indicated that they had been using this, although some more than others. A-R2 mentioned that she had been using it occasionally, but said that she would rather use the synchronising function more. A-R3 revealed that as she generated more data, her usage of this component increased correspondingly. The A-VRE-C stated that everybody in the group had been using it to store files on the system, by dragging and dropping them into the system from their devices. In Case Study B, B-R4 disclosed that she had been accessing her files via her iPad, and that she had been using the ‘My Files’ component for that.

- **My Activities (News)**

The answers from the respondents in both case studies revealed that none of them had really used the ‘My Activities (News)’ component. This could be because the size of each group was small and each member knew each other, which minimized the importance of using this component. The ‘Site Activities’ component was shown to be of more importance to the respondents.

- **Site Activities**

All of the respondents in both case studies were shown to have been engaging with this component. This component normally sent out reminders automatically via e-mail to the respondents in the VRE about any activity on the site. It also automatically sent out messages onto the dashboard of their sites. This component had thus been seen as a valuable tool to stay abreast of happenings within the sites. For example, B-R1 indicated that if a new article was uploaded onto the site, everyone was informed in a timely manner.

- **My Tasks (Workflow Function)**

The answers by respondents in Case Study A revealed that one respondent (A-R1) had been using it and six (A-R2, A-R3, A-R4, A-R5, A-VRE-C and A-VRE-M) had not. A-R1 indicated that she had used this when she was working on a manuscript, but also mentioned that after a period of time, people forgot what they had learned in the training sessions and reverted to using e-mails instead, because they were more familiar with it. She also mentioned that she had initiated a few of these workflows, but people in general did not respond. She was of the opinion that people had not responded, either because they had forgotten what they had learned, or because of a fear of the technology. A-R3 stated that she had not even known that it was there, which can be traced back to a lack of training, because she had joined the VRE at a later stage. A-R4 also mentioned that e-mail had been much easier, although not as secure. A-R5 knew about the component but said that having a shared folder where everyone just shared and uploaded had worked better for them. She also mentioned e-mail as working better for them, as the VRE had not been active (open for use) all the time, because one had to login to the system to use the component. The A-VRE-C stressed that it would only be used actively, when senior people requested researchers to use that.

Answers from respondents in Case Study B showed that this component had been used more extensively. Four of the respondents (B-R1, B-R2, B-R4, B-VRE-M) indicated that they had been using it, while two (B-R3 and B-L) indicated that they had not been using it. The usage of this component had been instigated by the B-VRE-M, who also acted as supervisor/promotor of these researchers. B-R1 and B-R2 mentioned that the supervisor would assign tasks to the researchers to attend to. They would then work on these and send the revised tasks back to the supervisor via the Workflow system. B-R3 indicated that she had not been using this yet, as she was still waiting for audiology equipment from Denmark and therefore had not been using the overall VRE actively yet. She mentioned that perhaps the writing of protocols could be done through this workflow system, but then it should be part of a policy that compelled one to use it. The B-L also indicated that she had not been using this component, because it is primarily a process between the supervisor and the researcher.

- **My Documents (Keeping Track Of Own Content)**

Four respondents (A-R1, A-R4, A-R5, and A-VRE-C) in Case Study A revealed that they had been using this component, and three (A-R2, A-R3, and A-VRE-M) revealed that they had not been using it.

In Case Study B, three of the respondents (B-R2, B-R4, and B-VRE-M) disclosed that they had been using the component and three (B-R1, B-R3, and B-L) indicated that they had not. B-R3 indicated that she had not really used it, as she did not have a large number of documents yet. B-R4 stressed the value of this component when she mentioned how easy it was to lose track, when one for example was working on the tenth draft. This component, according to her, could help hugely in tracking one's own content.

- **Shared Files (Files Everyone Has Access To)**

The interviews revealed that all the respondents in Case Study A had been using 'Shared Files'. A-R1 divulged that she had been using this on a daily basis, and she specifically mentioned a Liquid Nitrogen folder to which everyone had access. This folder had been used as a repository for everything related to Liquid Nitrogen. One could download files from there, work on them offline, and upload a new version of it there. The 'Site Activities' feed kept track of who had been using it. A-R2 mentioned that the A-VRE-C had created shared files where members of the group could access files that were common to the group. The A-VRE-C mentioned that some of these shared folders had been specific for specified groups of researchers in the VRE. Five of the respondents (B-R1, B-R2, B-R4, B-VRE-M, B-L) in Case Study B revealed that they had been using shared folders, while one respondent (B-R3) mentioned that she had not. The reason that B-R3 had not been using it was because she had not been using the system actively yet. Creating shared folders would seem to have been quite valuable for the sharing of files and articles that are common to each group.

- **People Finder**

The answers from respondents in both case studies revealed that none of them had used the 'People Finder' component. The reasons for this became clear through the answers received from A-VRE-C and B-R-2. A-VRE-C mentioned that she thought it would be more applicable if one had multiple sites of people, whereas B-R2 indicated that the group was so small that she did not need to use it to find people in the group.

- **Invite Users**

In Case Study A, one respondent (A-VRE-C) divulged that she had used this component, while six of the respondents (A-R1, A-R2, A-R3, A-R4, A-R5, and A-VRE-M) indicated that they had not. This component had generally been seen as an administrative responsibility, and the right to invite and add new users had been given to the A-VRE-C by the VRE-D as part of her role.

In Case Study B, the rights to invite and add new members to the group were given to the B-VRE-M, but due to technical problems they had experienced, the group had been relying on the VRE-D to assist them with this component. This was confirmed by B-R4 and the B-VRE-M in their replies to this question.

- **Discussions**

None of the respondents in Case Studies A and B had been using the 'Discussion' component. In Case Study B, B-R2 mentioned that they added comments at the bottom of documents, but did not really get into a discussion. B-L mentioned that she had gotten the impression that the group did not really discuss, but that each one had been doing their own thing. The reason for the non-use of this component could, as mentioned earlier, be due to the fact that the social aspect of the VRE had not been developed by the group by the time the interviews were conducted.

- **Document Library**

All the respondents from both case studies revealed that they had been using this component extensively to upload and download their documents and data. It seemed to be the most used component of the VRE. The A-VRE-C confirmed this when she indicated that the respondents used it the most. She also mentioned that they had common shared documents in the document library, which contained forms that needed to be completed, ethical documents that had to be accessible to everyone, and databases (folders) of samples that everyone could refer to. She also mentioned that each student had their own folder, which was only visible to them, and into which they could back-up their data.

The ‘Document Library’ on Alfresco had a number of sub-components, as discussed below.

- **Categories**

Two of the respondents (A-R1 and A-R3) in Case Study A indicated that they had used this component to organise and structure their files and folders, and five (A-R2, A-R4, A-R5, A-VRE-C and A-VRE-M) divulged that they had not made use of it. In Case Study B, five of the respondents (B-R1, B-R2, B-R4, B-VRE-M, B-L) mentioned that they had placed their files under specific categories, while one (B-R3) stated that she had not done this. B-R4 mentioned that she had used the ‘Categories’ component to structure her files and folders. The researcher of this study is of the opinion that the organising and structuring of files and folders under specific categories would make the files more accessible. Furthermore, the idea of placing files and folders under specific categories is a typical feature that can be found in social media such as blogs. Earlier, in the answers from respondents, it was mentioned that the social aspect of the VRE had not yet been developed much by either case study, which could be the reason why the majority had not been using the ‘Categories’ component as such.

- **Tags**

Two of the respondents (A-R2, A-R5) in Case Study A used tags and five of the respondents (A-R1, A-R3, A-R4, A-VRE-C, A-VRE-M) revealed that they had not been using tags. A-R1 indicated that the reason she had not been using tags is because the names of the files she uploaded onto Alfresco, were the same as on her computer’s hard drive. She stated that she knew where to go for what, and therefore did not need tags. A-R2 mentioned that she had only tagged once, when they had run a workshop. She had tagged a specific subject field, so that if people would search for it in future, they could find it easily. She indicated that, apart from that, she had not tagged any of her personal files. A-R5 stated that she had used it occasionally for the shared folder, when she uploaded or downloaded things there, and she foresaw that she might use it more in the future. A-VRE-C expressed that the members of the group had been using the VRE mostly for data backups, but did not think it had progressed beyond that. She admitted that this would become important for retrieval purposes in future.

In Case Study B, one respondent (B-R2) indicated that she had used tags, while five (B-R1, B-R3, B-R4, B-VRE-M, B-L) indicated that they had not. B-R1 disclosed that because she had arranged her files in specific named folders, she did not need tags to find anything. B-L pointed out that because she had only undergone a fraction of the training that the rest of the group had gone through, she had not been familiar with the component. It would have made things easier for her. The uptake of tags in both case studies were shown to be very low, which could have been, as the A-VRE-C stated, because members had been using the VRE more for backups than for future retrieval purposes.

- **Favourite**

None of the respondents in Case Study A had used this component, and only one respondent (B-R4) in Case Study B indicated that she had been using it. The A-VRE-C indicated that she had 'favoured' the group's site, but not the documents per se. B-R2 mentioned that the site had been so easy to access that it had been unnecessary to 'favourite' something specific. B-R4 stated that she had used it to sort out old things and to 'favourite' new things. The researcher of this study found that this component could be found in many social media platforms, but these two case studies, and especially Case Study A, had not yet developed the social features of the VRE much. It would seem they had been viewing this as a 'nice to have'.

- **Like**

The interviews revealed that none of the respondents in Case Study A or B had used the 'Like' component, for a similar reason as mentioned under 'Favourite', namely that the social features of these groups had not been developed much yet.

- **Comments**

Two of the respondents in Case Study A (A-R5, A-VRE-C) indicated that they had been using the ‘Comments’ feature and five of the respondents (A-R1, A-R2, A-R3, A-R4) stated that they had not. The A-VRE-C mentioned that she had been under the impression that the rest of the members of the VRE added comments to a new version of a file that they upload, because she had been doing this. The interviews, however, revealed that this had only been done by A-R5. The other respondents divulged that they had only used the VRE as a place to store and backup their data, and had not added any comments to a file in order to create context to the file. This could potentially in future create a problem when the number of files grows. It could also make it difficult for other researchers to understand the context in which a file was created, after a researcher has left the VRE.

In Case Study B, the usage of comments had been more evenly spread among members. Three of the respondents (B-R1, B-R2, and B-VRE-M) had been using comments and three (B-R3, B-R4, and B-L) had not. B-R1 stated that she had been using it to give information about the files, for example describing what type of documents they are (in other words, metadata or tags). The B-VRE-M admitted that she had been using the comments component only for a specific student. B-R4 admitted that she had only been reading other members’ comments, but had not added comments herself. In Case Study B, it seems that there had been a greater awareness of the value of adding comments, as can be deducted from the number of members adding comments. Adding of comments, as earlier mentioned, creates context to the files, and would make them more accessible in the future.

- **Share**

In Case Study A, three of the respondents (A-R1, A-R5, and A-VRE-C) stated that they had been using the ‘Share’ component, while four (A-R2, A-R3, A-R4, and A-VRE-M), indicated that they had not. A-R1 mentioned that she had used it to share files with some of her colleagues/peers. A-R5 divulged that they had tried it, but that it did not work. A-R5 then mentioned that the A-VRE-C created a shared folder where people could upload or access those files. She also indicated that she had occasionally created a new folder with its own rules about sharing documents.

In Case Study B, four of the respondents (B-R2, B-R4, B-VRE-M, and B-L) disclosed that they had been using the ‘Share’ component, while two (B-R1, B-R3) indicated that they had not. B-R1 mentioned that she had used other members’ shared files, but had not shared one of her files yet. B-R2 stated that she had shared documents with her supervisor (B-VRE-M), and B-R4 mentioned that she had shared articles with the rest of the members, while the B-L stated that she had only shared files once or twice.

- **Edit Properties**

Only one of the respondents in Case Study A, A-R5, indicated that she had sometimes edited her files’ properties. The rest of the respondents (A-R1, A-R2, A-R3, A-R4, A-VRE-C, A-VRE-M) indicated that they had never edited the properties of their files. In Case Study B, two of the respondents (B-R1 and B-R2) indicated that they had edited the properties of their files, but the rest (B-R3, B-R4, B-VRE-M, B-L) stated that they had never done that. B-R1 stated that she had renamed files, and B-R2 mentioned that she had edited the properties of her files on a regular basis. This is typically done to organise and structure files, and make them more accessible.

- **Edit Offline**

The ‘Edit Offline’ function (component) of Alfresco is one of the positive attributes of the Alfresco system, in that it allows the researchers to do their work offline, and then, when they go online, the updated version of the file synchronises to the version that is on file and updates it. In Case Study A, four of the respondents (A-R1, A-R3, A-R4, and A-VRE-C) mentioned that they had been using the function, while three (A-R2, A-R5, and A-VRE-M) indicated that they had not been using it. A-R1 revealed that she had used it often, while A-R3 stated that it had been problematic to update the Liquid Nitrogen Database Folder online, and that they had been editing it offline, and then uploaded it again afterwards. A-R4 divulged that he had been using the function, but not as frequently, because he had to login to Alfresco to access or synchronise files. He felt that this should be an automatic process. He also had been using Google Drive, Dropbox, and OneDrive to store files. A-R5 mentioned that she had not even been aware that there was such a function that she could use. This was consistent with the lack of training about the VRE. The A-VRE-C indicated that the group had been using this function regularly. According to her, they had tried to use Google Docs for editing, but that there had been a problem to get the two systems (Google Drive and Alfresco) to interoperate. As an alternative, the group had been able to download a file, work on it offline and then upload a new version again.

In Case Study B, one of the respondents (B-R2) disclosed that she had been using the function, and five (B-R1, B-R3, B-R4, B-VRE-M, and B-L) indicated that they had not been using this function. B-R3 mentioned that she had not used this function as she had not been using the site yet. This is in line with her answer earlier, where she indicated that she had been waiting on instrumentation from Denmark. It would seem that the majority of respondents in Case Study B had a preference to do everything online, as mentioned by B-R4, or they had not thought to use the ‘Edit Offline’ function, as mentioned by the B-L.

It is interesting to see the difference between the two case studies. More respondents in Case Study A had been using the offline editing function, which could be because of their work in the laboratory, where they collected data using various instruments and machines that were not linked to Alfresco. Data were then stored and edited offline and later uploaded or synchronised to the live version of Alfresco. In Case Study B, more respondents had been using the live version than the offline function, because their tools had already been plugged into Alfresco, for example the survey tool.

- **Dublin Core Metadata Template**

Alfresco has a Dublin Core Metadata Template that is already included in the platform, but the interviews revealed that none of the respondents in either of the case studies had made use of this component. Two of the respondents (A-R4 and

A-VRE-C) admitted that this was something that should be talked about and used, but it would seem that the majority of the members in these case studies did not understand the necessity or value of adding metadata to their documents and data. In addition to this, not everyone understood what is meant with metadata, as was revealed by the answer received from B-R2, when she indicated that she did not know what the term meant. Another respondent, B-R4, felt that it had been too cumbersome to complete the Dublin Core Metadata Template. Researchers were often in a hurry and it would seem that they just needed a secure space to save their data. This could be a reason why most of the respondents had not completed the Dublin Core Metadata Template. This is in line with the discussion under ‘Tags’, where it was mentioned that the student researchers only used the VRE as a place to backup and store their data. ‘Tags’, which could be described as a type of metadata, could be one way of ensuring the easy retrieval of data, and the Dublin Core Metadata Template could provide even more fields that could ensure the successful retrieval of data in the future. The fact that none of the members of either of the case studies used this component, and very few used the ‘Tags’ component, is a cause of concern. These components are important for understanding the context and provenance of data, and are essential for the retrievability of data in the future.

- **Manage Permissions**

All the respondents in Case Study A indicated that the A-VRE-C and the VRE-D managed the permissions in the VRE. A-R1 mentioned that if she wanted to share something and then experienced problems, she would send an e-mail to the A-VRE-C, who would then contact the VRE-D. He would then change the permissions. The A-VRE-C stated that she had been using that often. In Case Study B, all the respondents indicated that they had not used this themselves. B-R2 indicated that the VRE-D had been handling the permissions, and if any of them needed to add someone to the group or to a specific document or file, they would contact the VRE-D. The B-VRE-M mentioned that she held the rights to do that, but hadn’t done that yet. She indicated that she had nevertheless been using e-mail extensively, but also realised that as a group, they needed to get used to implementing the workflow on the system.

- **Upload New Version**

In 7.2.1.8, in the formative evaluation, it was mentioned that the Moodle platform was replaced with the Alfresco platform, because the members of Case Study A had indicated that they needed a versioning function/component, and that Alfresco could provide this component, which Moodle could not.

During the interviews, it was revealed that three of the respondents from Case Study A (A-R1, A-R2, and A-VRE-C) had been using the ‘Upload New Version’ component, and three (A-R3, A-R4, A-R5) had not been uploading a new version. The fact that only three used this component, at the time of the interviews, seemed to contradict the group’s initial request, during the formative evaluation, for a system that could do versioning. The individual answers received from the non-users, however, shed some light on the reasons for this. A-R3 and A-R5 had not been aware that the system automatically created a new version of a file when they made changes to a file and uploaded it again, and that the system kept all the previous versions of a file. This could be because both of these respondents missed out on the training that the rest of the respondents had undergone. As indicated earlier, A-R3 had joined the VRE at a later stage, and A-R5 could not attend the hands-on training session. The interview with A-R4 revealed that he didn’t trust the system to keep all his versions, and he expressed his anxiety that his files might get corrupted. For this reason, he had been giving each new version of a file, a new file name.

Three of the respondents (B-R1, B-R2, B-R4) in Case Study B divulged that they had been using the ‘Upload a New Version’ function, but three (B-R3, B-VRE-M, B-L) stated that they had not been using it. This is in line with an earlier remark by B-R3 that she had not used the system yet. The B-VRE-M revealed that she had been using the ‘My Tasks (Workflow)’ component to send revised documents to her students and vice versa, but had not been uploading new versions of files yet. She did indicate that she was aware of the value of this function and would probably be using it in the future.

- **Download Function**

The interviews with the respondents revealed that all of the respondents in Case Study A had been using this function, but in Case Study B, only four of the respondents (B-R1, B-R2-B-R4, B-VRE-M) had been using this function, while two (B-R3 and B-L) had not been using it. As mentioned earlier, B-R3 had not been using the system yet, while B-L indicated that she had been uploading documents onto a student's folder only, but had not downloaded anything. This probably had to do with her role as librarian to help provide information.

- **Instrument Backups**

In both case studies, the respondents indicated that they had not used the 'Instrument Backups' component of the VRE. Reasons provided were varied. A-R1 mentioned that the instruments they had been using did not allow connectivity to the Internet, and a vast number of instruments did not have automatic backups. She indicated further that they usually plugged in a hard drive to an instrument, downloaded everything, and then did mass backups to Alfresco manually from the hard drive. A-R2 stated that her work did not generate data on instruments. A-R4 mentioned that the group did not have any instrument linked to the VRE, which confirmed what A-R1 had said. A-R5 disclosed that she had been giving it to the A-VRE-C to do, and the A-VRE-C revealed that she had been backing up instruments and then uploaded these backups onto Alfresco. The answers from Case Study A revealed a need for instruments to be linked to the VRE so that their data could be uploaded and archived on the VRE platform. The answers received from respondents in Case Study B revealed that not many of the researchers had been using instruments to do their research. B-R1 mentioned that she had uploaded the backups of instruments that she used, under her own profile on Alfresco, while the B-VRE-M felt that it was not relevant to them at the time of the interview, but might be of importance to them in the future. The possibility for instruments to be plugged into a VRE is mentioned in 4.5.2.1 and 5.4.

- **Software Backups**

The interviews with respondents revealed that none of the respondents in Case Study A had been using the software backups component. In contrast, one of the respondents (B-R4) in Case Study B indicated that she had been using the component; however, five (B-R1, B-R2, B-R3, B-VRE-M, and B-L) stated that they had not been using it. A-R1 divulged that the software she had been using had an automatic backup on her laptop computer, with the result that she had not needed to back it up on the VRE as well. The researcher of this study identified this as a risk, if something happened to her laptop computer. A-R5 stated that she had not been aware of the component, and mentioned that it would be a valuable component to use to back-up her software. She further mentioned a statistics programme that she had recently purchased, and would like to back that up on the VRE. B-R1 also indicated that she had not been aware of the component. It could have been something she had given a miss during the training sessions. B-R4 on the other hand mentioned that she had been using this component to back-up the software and programmes of the 3D capturing system that she had been using for her doctoral research, on Alfresco.

- **Survey or Questionnaire Tool**

All the respondents in Case Study A indicated that they had not been using this tool. This had been expected, as the nature of their field of research was to use natural science-oriented data and laboratory/experimental methods, and not surveys or questionnaires. In Case Study B, the picture looked similar, but slightly different. Even though none of the respondents had used it yet, B-R2 mentioned that she would be using it in the next phase of her research study. B-R3 also mentioned that she would be using it in her study. The B-VRE-M revealed that they had a number of projects in the group that this tool would be relevant for, and planned to run a pilot study on the survey tool.

- **Publishing Function**

This function allows researchers to publish their data or findings on social media platforms. None of the respondents in Case Study A or B, however, had used this component. In Case Study A, A-R1 mentioned that they had not been allowed to put any of their research activities onto social media, because of intellectual property issues, which could be the reason why none of the other respondents had been using it. A-R3 indicated that she had not even been aware of the function, while the A-VRE-C stressed that they had not yet used the social aspect of Alfresco. In Case Study B, B-R1 revealed that she had not used it, because she hadn't reached the publishing stage of her research yet, which could also be the case with the other respondents.

- **Mobile Syncing With Alfresco**

Alfresco refers to synchronisation as syncing (see also 7.2.1.5 and 7.2.1.8). This functionality requires that the members of the VRE download an Alfresco app onto their mobile phones or tablets. The app can then be used to synchronise their files (on their mobile devices) with the files on Alfresco. One of the respondents (A-VRE-C) in Case Study A revealed that she had used the mobile syncing component with Alfresco, but six (A-R1, A-R2, A-R3, A-R4, A-R5, A-VRE-M), indicated that they had not used it. A-R1 mentioned that she had not even been aware of the mobile app for Alfresco, and also revealed that she had not stored all her documents on her phone, which could inhibit her from using this component. However, she stated that it might help for other things. A-R3 had also not been aware of the component, which is again a reflection of the lack of training. A-R2 and A-R4 indicated that they did have the app, but had not used it to synchronise to Alfresco. Desktop syncing seemed to be more preferable. The A-VRE-C revealed that she had used the mobile syncing component, especially when she had been in meetings. A-R5 divulged that she has been experiencing problems to get the app to work on her phone, which meant that she could not use it.

In Case Study B, two of the respondents (B-R4 and B-L) mentioned that they had been using this component, and four (B-R1, B-R2, B-R3, B-VRE-M) stated that they had not been using it. B-R4 mentioned that it had been working very well for her, and that she had been using her iPad tablet to do the syncing. B-L indicated that she had been using her tablet to do that, especially when asked for something during times away from her office. B-R2 revealed that she had mostly been using her laptop to synchronise and not really her mobile phone.

- **Desktop Syncing With Alfresco**

Five of the respondents (A-R1, A-R2, A-R3, A-R4, and A-VRE-C) in Case Study A stated that they had been using the ‘Desktop Syncing’ component, and two (A-R5, A-VRE-M) revealed that they had not. A-R1 mentioned that she had been encountering problems with this component. She indicated that when she synchronises a new version of a file to the VRE, the system did not update the new version. At the time of the interview, she and the A-VRE-C were busy sorting out the problem with the VRE-D. A-R5 knew about the functionality of desktop syncing, but had not been using it. She could not really give a reason, but it probably comes back to preference. She had been backing up her data on external hard drives, because she found that easier for uploading and downloading of data. Then, when a project had been completed, she uploaded it onto the VRE, which she described as an archive system. She indicated that she did not find the VRE user-friendly, which could be because she had not attended the hands-on training session. The A-VRE-C reported that they had been experiencing a few issues with synchronising, but that they would be solving it together with the VRE-D. According to her, synchronising was essential to avoid duplication when backing up things.

In Case Study B, four of the respondents (B-R2, B-R4, B-VRE-M, and B-L) indicated that they had been using the ‘Desktop Syncing’ component, and two (B-R1 and B-R3) indicated that they had not been using it. B-R2 and B-R4 mentioned that they had been using the component from their laptop computers because all their documents were on the hard drives. This is where they did their typing either in MS Word or MS Excel. The reason B-R1 had not been using the component seemed to be a matter of preference. B-R3 revealed that she had not really started working on the system and therefore had not used this component as yet.

(9) Do the tools available in the VRE meet all your expectations? If not, why not?

In Case Study A, A-R1 felt that the components or tools in the VRE met their current needs, but that there had been many of the functionalities that they had not used. A-R2 again mentioned the problem with synchronising, and A-R3 felt that she was unqualified to answer the question, because she had not tested all the functionalities of the VRE as yet. A-R4 stressed that it was difficult to measure, because he had not used it routinely.

He had used many other tools outside the VRE, and was not convinced that the VRE was the better tool for him. He felt that it needed to be part of his daily routine. He mentioned further that the system looked a bit bland and should be constantly upgraded or further developed. He described the VRE as functional, but not what he would have liked to see. A-R5 described it as a great tool or safety net for the backup of data, where one could store one’s data for a long time and also make one’s data accessible. The system, according to her, had not been the best system for interaction among members of the group. This could be because the social aspect of the group had not been developed as yet, as mentioned earlier by the A-VRE-C. The A-VRE-C stressed that the tools of the VRE met the requirements of the group, at the time of the interview, which was the backup of their documents. She indicated that the system could be a bit more ‘smooth’ or user-friendly. The A-VRE-M indicated that he saw the VRE as the start of a wider research process.

In Case Study B, B-R1 mentioned that she had used the VRE as needed. It had also been easier for the B-VRE-M to upload a video for her there, which she was able to access. This respondent (B-R1) stays geographically far away from the University, in the countryside, and the VRE had made things much easier for her. B-R2 indicated that she did not stay in South Africa, and the VRE had made it much easier not only for her, but also for her supervisor. She stressed that her supervisor as well as other participants in her study could, through the VRE, have access to documents that they needed, and that the supervisor could track her progress. She found the functionality of sending others in the group a reminder via e-mail, when one has uploaded something, as a great selling point. She felt, however, that the system lacked a link to a referencing system, for example EndNote or RefWorks, and she would have liked to see that added. B-R3 couldn’t really give an answer, as she had not used the system much yet. She nevertheless had some analysis software that she was going to use for systematic literature reviews, and indicated that she would like to see these integrated with the VRE, if possible. B-R4 mentioned that she was very satisfied with the functions of the system that they had been using, but were of the opinion that they had been underutilising the system. The B-VRE-M stated that she needed to ‘immigrate’ into the system so that she could use the system more with her students. She stated that she would like to have the Qualisys Track Manager System (an electronic movement analysis system) added to the VRE, because it would be important for their publications.

She also mentioned another project that might have software programmes that would need to be added nearer to the data analysis stage. The B-L stated that she was satisfied with what they had been using, but she would prefer to experiment more with other functionalities of the system to see what the possibilities could be.

(10) Do you have someone from your research group / department that acts as champion / facilitator for the group? What is his / her designation / title in the group?

In Case Study A, all the respondents indicated that the A-VRE-C had been the facilitator / champion of the group. Two respondents (A-R1 and A-R2) mentioned that the A-VRE-M sometimes acted as facilitator.

In Case Study B, the respondents disclosed that the B-VRE-M had been the facilitator / champion of the group.

Section B: Questions on RDM

(11) What would you describe as research data? Why do you see some data as not research data?

As mentioned in 6.3, the purpose with this question was to gauge the respondents' knowledge about research data, and what they did not consider as research data. This links up with the distinction that was made between research data, referencing data, funding data, collaboration data, and administrative data in 4.2.2. In Case Study A, A-R1 described research data as the inputs and outputs of one's work. Inputs were described as articles that one has read, and documents one has written. Outputs were seen as results from instruments. These inputs and outputs then together constitute a manuscript. This description of what constitutes research data differed from the definitions given in 4.2.2, where outputs, such as documents that one has written, are not included as research data, unless used as a source document for further research. A-R2 saw research data as anything that one generates in the course of one's experiments. A-R3 described research data as all the raw data that one collects, all of one's processed data, one's digitised lab book, as well as all the research articles that one had used. Her description was in line with the definition of research data in 4.2.2. A-R4 defined research data as anything from a raw data file that came off an instrument, to pre-analysed (cleansed data), analysed data, statistical reports, and even writing a paper, article or thesis. He also included the planning of an experiment in his description of research data. The researcher of this study, however, classified this as administrative data in 4.2.2. According to A-R4's definition, it seemed all of one's inputs and outputs could be described as research data. This is, however, not in line with the definitions found in literature (see 4.2.2), where outputs such as a paper, an article that one produces, or a thesis that one writes, are not regarded as research data unless used by another researcher as data.

A-R5 depicted research data as data generated through research activities, using different research tools, for example a microscope image, or data in a Microsoft Xcel spreadsheet that were generated through another system. This was more in line with the definition of research data as found in 4.2.2. The A-VRE-C described research data as everything they generated. This would include raw data that they generate (e.g. Microsoft Xcel files, photos, Flow Cytometry data, and Kaluza files), as well as processed data (e.g. statistical files and Xcel files), which is in line with the definition of research data in 4.2.1. She did, however, regard an article or a thesis that flowed from this research as the end product, but felt that these should be treated in the same manner as research data, in terms of storage. The A-VRE-M described research data in terms of the types of data that they generated or worked with. He mentioned that they had a big quantity of data that came from a Flow Cytometer, half a dozen really big databases that were mostly Microsoft Xcel-based, and a big amount of genomics data. According to him, the VRE also had to be able to handle data from specific national projects that they were involved in.

In Case Study B, B-R1 saw research data as any data that could serve as a resource for one's research, or contributed to the research. This, according to her, could include anything from class notes, protocol development, forms, objective measurements, to articles. This differs, however, from the description given in 4.2.2, where protocol development and forms were seen as part of Administrative Data. According to B-R2, research data started from the beginning, where one's protocol commenced. This included all information leaflets and all ethical approvals. This description deviates from the researcher of this study's distinction of different data types in 4.2.2, where ethics approvals were described as part of administrative data, and not research data. In the next phase of the research cycle, which is the literature search, B-R2 indicated that she would include all the articles she had consulted and then upload these, which is line with the definition of research data in 4.2.2, but then she would also include all the search strategies that she had undertaken, which could rather be described as referencing data. Following this, in the next phase of her study, she mentioned that she would include all field notes and all the questionnaires that she had used. She focused on quantitative and qualitative research, but for the qualitative part, she needed to make sure of the provenance of the data. In other words, she needed to document everything so that anyone could scrutinise her data and see how it had been done.

B-R3 described research data from a quantitative perspective. She saw research data as the scores from her different outcomes-measures. She described these as numbers, pictures (images and photographs), and video clips. She also mentioned that systematic reviews, literature reviews, questionnaires, theses and articles could be data, which is very much in line with the discussion in 4.2.2 on numeric, visual, and textual data as types of research data that could be used for analysis. B-R4 depicted research data as all the literature that they had used, everything that they had collected through fieldwork and discussions, which are in line with 4.2.2. However, they also added things that they had written, which sounded more like outputs than research data. The B-VRE-M mentioned that what could be seen as research data, would be dependent on the type of project that one did. She described research data as observations that one made, either through technology, or that one had recorded according to certain criteria and guidelines on forms and tick lists. Research data, according to her, could therefore consist of documents, graphs and video material (recorded by either video recorders or specialised electronic movement analysis programmes), photos and specialised data that they had used in their projects, which is in line with 4.2.1. A qualitative approach would, according to her, also include protocol development, protocol defence, the ethical process, transcriptions of interviews and focus group interviews, surveys, history and processes of literature search strategies, selecting appropriate resources, the coding process, a conceptual framework, and final reports and publications. In other words, she saw everything they do in the research process, as research data. Her description deviated from 4.2.1, where protocol development, protocol defence, and the ethical process were highlighted as administrative data. Transcriptions of interviews, focus group interviews, selecting appropriate resources, and the coding process can be seen as research data, while the conceptual framework, final reports and publications can be described as outputs. The researcher of this study tried to probe whether the respondent would classify something like ethical clearance as a different type of data, for example administrative data (see 4.2.1). The B-VRE-M agreed that it could be seen as administrative data, but stated that it could also be seen as part of research data. She felt that it was important to capture the whole process of getting ethical clearance, which then culminates in the ethical process. The reason for this is that their ethical clearance needed to be stated in all their published documents. She also saw the clinical trials that they had to register at the National Clinical Trials Registry, or at the Pan-African Clinical Trials Registry, as part of the research data that needed to be stored. The researcher of this study included this type of data under administrative data (see 4.2.1), and not under research data, but agreed with B-VRE-M that this type of data needed to be stored in a VRE.

The B-L described research data as the information that had been gathered to apply to the product that would be presented or written-up, and the VRE-D described research data as data that had been gathered from a set of raw data, and which is in the process of being processed and analysed, to either confirm or reject the researcher's objective. This definition of research data is in line with the researcher of this study's definition of research data in 4.2.1.

(12) How would you describe the concept 'Research Data Management'?

As mentioned in 6.3, this question was asked to determine what the respondents understood under the RDM concept, as their answers would have a direct impact on answers to further questions. In Case Study A, A-R1 equated RDM with backup, which sounded a bit narrow as RDM is so much more than just backups. She did, however, elaborate a bit more on RDM. According to her, it was not only about saving data in the right format so that it is easily accessible, but also about ensuring that the data were not lost. RDM, according to her, could also include other people, thereby ensuring that one had the right tools to manage the data. A-R2 felt that the concept of RDM didn't really mean anything to her, but was of the opinion that it was something that group leaders think about. She admitted that she managed her own data, but stated that it was not an overly complicated process. She had been doing record keeping, note taking, and had seen to it that she was consistent with things such as file naming, file storage, and backing up of her data. RDM, according to her, was a much more serious process for the research group leader, because having access to the data generated by the people in one's group was important in terms of publishing, but also in terms of making decisions for what experiments or projects to embark upon next. A-R3 saw RDM as the proper storage of data, cataloguing of data, saving of data, and findability of data. A-R4 recommended that the University should have a more rigid structure, for example a data management office, in place to look after the data that are generated within the University. The reason he gave for this was that public funds were used to do most of their research. He then mentioned examples of two different types of data that should be managed (looked after responsibly): raw data generated from equipment or machines, and processed data. A-R5 saw RDM as occurring in two areas, namely maintenance or storage of existing data, and managing accessibility to research data, so that it would be easy to track and find things again, and also to see who was using what. The A-VRE-C described RDM as storing data in a way that would make it understandable for future usage. The A-VRE-M mentioned that researchers could have expensive hardware and great ideas, but if the data were not managed properly, it would be useless. He suggested that the management of data at the University should not be done in a way that is top-heavy. In other words, the method of doing RDM should not be forced down on researchers by University management or anyone else. The RDM process at the University should make it easier for researchers to do their research.

Synthesizing the responses received from Case Study A, the majority of respondents in Case Study A focused on the storage and accessibility of data, and this correlated with the first part of Texas A & M University Libraries' (n.d.) definition as given in 4.2.3.6, namely "storage, access and preservation of data produced from a given investigation." The respondents failed, however, to mention the preservation of data, although A-R1 did mention that one must make sure that data are not lost, which could be deducted as referring to preservation. None of the respondents mentioned the research data lifecycle.

In Case Study B, B-R1 described RDM as certain steps that one could follow to handle storage of data during a research project. This would include documenting the process and archiving the document (data) in a safe place. B-R2 indicated that RDM included access to data, anywhere, anytime, and without limit, as well as the backup and saving of data. B-R3 saw RDM as the collection of all data, putting it on something like a spreadsheet and then uploading of data onto a system such as a VRE. B-R4 described RDM as all of one's literature searches, search strategies, and data saved and stored on a system such as a VRE. The B-VRE-M saw RDM in terms of a research data lifecycle (see 4.5 for a discussion on research data lifecycle). The way it was done, according to her, was determined by the nature of the research project. The process starts when one decides on a topic for one's research, continuing one's observations, interviews, and focus group discussions and proceeding up to the publication of one's research. Throughout the process, one has to keep track of the data gathered. RDM, according to her, also enabled correlation between the raw data, conclusions and publications. She was furthermore of the opinion that RDM enabled the research leader to integrate data generated by different projects. A proper RDM system, as stated by her, would also assist researchers in storing data for 15 years as required by the University.

The B-L described RDM as being a vague concept to her, but then delineated it as ensuring the availability of one's data, so that one's data could be utilised by others, in different environments. The VRE-D saw RDM in terms of the research data lifecycle. According to him, the RDM process starts when the researcher selected his topic after doing a bit of research beforehand, he/she then creates a DMP indicating what is going to happen to the data, where it is going to be stored, how it is going to be stored, and for how long it will be stored, taking into account the University and funders' guidelines. This correlates with Penn State University Libraries' (2014) definition of RDM as mentioned in 4.2.2.5: "How data is managed depends on the types of data involved, how data is collected and stored, and how it is used - throughout the research lifecycle." Texas A & M University Libraries (n.d.) further mentions the aspect of planning in their definition in 4.2.2.5, which also correlates with the idea of creating a DMP, as mentioned above. The VRE-D further mentioned that a DMP should also include information on the types of data that will be generated, all the way through to the publishing and dissemination phase, and also to preserving and re-using his/her data. The VRE-D felt that there was a lack in this area of data management planning, and that the majority of researchers had not been doing this.

Synthesizing the responses received from respondents in Case Study B, the majority of respondents (B-R1, B-R2, B-R3, and B-R4) described RDM in terms of storage of data, which is only one aspect of RDM. The B-VRE-M and the VRE-D, however, described RDM in terms of the RDM lifecycle, which correlates with the definitions given in 4.2.3.1 as well as the discussion of the RDM cycle in 4.5 of this study.

(13) To what extent would you say that research data could be managed through a VRE?

In Case Study A, A-R1 stated that it was definitely possible that data could be managed through a VRE. She indicated that they had used the Alfresco system and that it had been an excellent system to use for that. She also mentioned that the system allowed a researcher to stratify his/her work and to save it easily. A-R2 indicated that she had been mainly interacting with the VRE system (Alfresco) as a way of backing-up her data, and saw it as part of managing her research data. She also mentioned that it was possible to add metadata to one's data, which made it easier for the VRE Manager (A-VRE-M) to keep track of what one has been doing, while, in addition, enabling easy access to the data. A-R3 mentioned that the system enabled RDM quite nicely, and then continued to talk about the ability to save one's files, tag one's files, share one's files, and to protect one's files so that only certain people could see it. A-R4 suggested that the instruments they used should be linked to the system so that a data file that came directly off an instrument was automatically protected and backed-up. He also wanted it to be accessible from wherever he was. He would furthermore like to have a space or niche on the system for everything they did during the research process. A-R5 indicated that she was of the opinion that it was a great archiving tool, and very useful in sharing large data files with others, even if they were not part of the site. She then gave an example of a student in Geneva, Switzerland, with whom she had been sharing files. She was also impressed with the versioning function. According to her, the system, in addition, had a good monitoring function that kept track of who had access to certain data, who downloaded a file, and who uploaded a file, etc. The A-VRE-C stressed that it had been important to use the VRE every day. She was of the opinion that a system like that was a necessity for them, especially because they worked in laboratories. This system had been quite valuable to her in terms of storing their data responsibly and correctly. The A-VRE-M stated that the management of research data and the storage of it had been the major drivers for the establishment of the VRE. He stressed further that the emphasis should be on quality and not quantity, and that the success of a project lies rather in how it is managed, and how one can access it.

In Case Study B, B-R1 was of the opinion that research data could definitely be managed through a VRE. She then presented an example. As part of her research, she had made recordings of interviews and had been uploading these recordings onto Alfresco as she went along. She had been storing it under specific dates and times. B-R2 indicated that she was of the view that RDM could be done to the fullest extent through a VRE. She did mention, however, that uploading of big files (e.g. video files) was problematic, especially because of the Internet connection in Ghana, where she had been staying at the time. She indicated further that she thought the VRE was the way to go for future students, because one could have access to one's files anywhere. B-R3 stated that she felt unqualified to answer this question because she had not been using the system much at that point. B-R4 mentioned that the VRE had made things much easier for her, especially the storage of data. The system had given her peace of mind, in the knowledge that everything that she had uploaded was safe and accessible. In the VRE she also had someone that looked after the data with her. The B-VRE-M viewed Alfresco as the anchor and basis of all, and that everything in it was related to the research projects and processes. She also called it a monitoring and storage system. Alfresco, according to her, had been the ideal system for RDM. She furthermore mentioned that the VRE could be used to upload and centrally store everything. In addition, she stated that they were increasingly expected to provide proof of the research process and when they published, and that the VRE was becoming an essential tool to enable that. The B-L did not elaborate much, but said that it would be a substantial way of managing research data. The VRE-D stated that the VRE provided a central place for the researcher's data. The VRE gave the researcher a sense that when he/she logs onto the system, his/her data would be taken care off. The researcher did not necessarily see the infrastructure that his data was stored on. It could be stored in different plugins or servers, for example. The VRE, irrespective of file types, or various plugins, was a central place where data were located, not necessarily stored. That was where the research environment played a crucial role for the researcher. The VRE created a controlled personal environment for the researcher to do his/her work. The VRE pulled everything together in a specific technological framework or space.

As mentioned in 6.3, this question touched on the core of the research problem that this study aims to address. The answers received from respondents showed that they were using the VRE primarily for the backup function of RDM, for data sharing within the group, and the workflow function within the VRE, but not for all components of RDM. The majority of respondents indicated that they were using the VRE to save, back-up, store or archive their data, which were in line with Carusi and Reimer's (2010: 18-19) viewpoint, mentioned in 5.3, namely that VREs can provide an easy to use platform where researchers can secure the short-term storage of their data, and also afford them the means to keep control of their work. It was also in line with Neuroth, Lohmeyer and Smith's (2011: 225) viewpoint, that a VRE could provide researchers with a safe place where they can save their data directly.

Another aspect of RDM that was mentioned by the respondents was that of data sharing among members of the VRE. This corresponded with what was said by a number of authors as cited in 5.3. Carusi and Reimer (2010: 19) mentions the sharing of data with peers, and JISC (2006) as well as Carusi and Reimer (2010: 20) indicate that VREs provided the possibility to "share data and collaborate," while Filetti and Gnauck (2011: 237) saw data sharing as a key element in a VRE.

The idea of sharing data with others who are geographically dispersed were also mentioned by two of the respondents, and this was in line with Carusi and Reimer (2010: 18) and Neuroth, Lohmeyer and Smith (2011: 223, 230), cited in 5.3, who indicate that VREs could provide access to data as well as to co-researchers that are geographically spread out.

Access to data was another characteristic of VREs that was mentioned by the respondents. This corresponded with Carusi and Reimer's (2010: 13) view, mentioned in 5.3, that VREs presented a technological framework that provided access to data, tools and services.

Another aspect mentioned by the respondents was securing data safely, which corresponded with Brown (2013) and Robertson Library's (n.d.) views, cited in 5.3, that VREs provided the means and effective ways for securing data collaboratively.

Other aspects of RDM mentioned by the respondents, that VREs could handle, included adding of metadata to data (based on Dublin Core), tagging of data files, versioning of data files, and monitoring of data. These aspects were not found in literature dealing with VREs.

The aspect of collaboration that was mentioned by many of the authors in the literature (Brown, 2013; Carusi and Reimer, 2010: 10, 13, 19-20; JISC, 2006; Robertson Library, n.d.) was not mentioned by the respondents, and reflected the fact that the social aspect of the VREs had not been fully developed at the time of the interviews.

(14) How did you manage your research data before becoming part of the VRE project?

The answers to this question as received from respondents are listed in Table 7.2 below. In the left-hand column, the devices or tools that had been used to manage their data are listed, and in the middle column, the actions that were taken with these devices. These were matched in the right-hand column to the respondents who indicated that they had used them.

Table 7.2: Managing Research Data Before Becoming Part Of The VRE

Device	Action	Respondent
NAS (Network Attached Storage)	Storage	A-R1 used this at home. Her laptop synchronised everything she had done on her laptop to the NAS.
Cloud (e.g. Dropbox and Google Drive)	Storage Storage	A-R1 A-R3
External hard drive	Saved everything to it at the end of each week. Storage Storage	A-R1 saved everything to a hard drive at the end of the week. A-R2 A-R3 had two external hard drives.

	Storage	A-R5
	Storage	A-VRE-C
	Storage	B-R1
PC / Laptop hard drive	Stored thesis, pictures and results of studies on it	A-R2
	Storage	A-R3
	Storage of images	A-VRE-C
	Storage	B-R1
	Storage	B-R2
	Storage	B-R4
	Storage	B-VRE-M
Paper lab book	Recorded everything in the book.	A-R2
	Recorded everything in the book.	A-R3
	Recorded everything in the book.	A-VRE-M
CD ROM	Back-ups	A-R4
	Storage	A-R5
Paper-based files	Records, lab work	A-VRE-C
	Storage	A-VRE-M
	Storage	B-R2
	Files were in Excel and stored on paper.	B-R4
Stiffy drives	Storage	B-VRE-M
Floppy drives	Storage	B-VRE-M
Flash drives / Memory sticks	Storage	B-VRE-M
	Storage	B-L

The respondents in both case studies as shown in Table 7.2 indicated that they had used a variety of tools to store their data. The majority (A-R1, A-R2, A-R3, A-R5, A-VRE-C, B-R1, B-R2, B-R4, and B-VRE-M) had been using external hard drives and the hard drives of their PCs or laptops. Other devices mentioned in Table 7.2 included NAS (used by A-R1), CD ROMS (used by A-R4 and A-R5), stiffy drives (used by B-VRE-M), floppy drives (used by B-VRE-M), flash drives/memory sticks (used by B-VRE-M and B-L), paper lab books (used by A-R2, A-R3, A-VRE-M), paper-based files (used by A-VRE-C, A-VRE-M, B-R2, and B-R4), and cloud services such as Google Drive and DropBox (used by A-R1, and A-R3). These respondents indicated that they were still using some of these devices (NAS, external hard drives, the hard drives of their PCs or laptops, flash drives, paper lab books) and cloud services (e.g. Google Drive and DropBox), in addition to the VRE. All the respondents indicated that the VRE had given them a secure and safe environment to store and archive their files, and also enabled accessibility to their files, anywhere, anytime. This continued use of other devices to store copies of their data could be a reflection of research practices that had been in place over a long period of time before the VRE existed and were therefore familiar to the respondents. However, it could also be an indication that the respondents did not trust the VRE system fully yet.

(15) Are you able to do the following RDM related tasks / actions within the VRE: (please explain how it is done and if not, why not)

(a) Create / capture data, using:

(i) <u>Major Instruments</u>

In 5.4, the researcher of this study identified data capture/collection tools as RDM tools that can be added to a VRE platform. Examples of these were also discussed in 4.5.2, namely observations, textual or visual analysis, interviews, focus group interviews, surveys, tracking, experiments (using laboratory instruments), sensor instruments, case studies, literature reviews, questionnaires, etc. This question, however, only focused on data that are captured/generated through any kinds of instruments, and whether this can be done within a VRE.

All the respondents in Case Study A and B indicated that they had not been using the VRE to access major instruments. A-R1, A-R2, and B-R1 stated that they thought it could be done, but at the time of the interviews, no instruments had been linked to the VRE. A-R3 expressed her doubts that the VRE would be able to process the huge amount of data they were working with. The VRE, according to her, would need a lot of processing power to be able to process the data. A-VRE-C indicated that she had been gathering information/data through some instruments and had then uploaded and stored that on the VRE manually. B-R2 mentioned that she and her supervisor (B-VRE-M) had been using the camera function and the dictaphone app on their iPads to record data. These files were then manually uploaded onto Alfresco. The B-VRE-M confirmed this and indicated that, at the time of the interview, they had been uploading data from instruments manually, but were planning to have an automated (linked) functionality in the future. The B-L expressed her concern that such a linked functionality might be expensive. The VRE-D also confirmed that data from instruments had been uploaded manually. He further mentioned that some of the data had been uploaded via the drag-and-drop functionality and via the synchronise function, directly onto the VRE. The answers received from respondents revealed that there existed a need to plug-in their instruments to the VRE, so that they could access it through the VRE, and that they could upload/capture data automatically into the VRE from various instruments.

(ii) Simulations

Simulations were identified as a pluggable component in the 'other pluggable software components' layer of the proposed conceptual model in 3.5.7.3 and 5.4. Although this was not identified as a RDM tool per se, it is a tool that is, for example, used to do a simulation of an experiment and in the process, generates data.

All the respondents in Case Study A and B indicated that up to the time of the interviews, it had not been possible to run simulations through the VRE. The B-VRE-M indicated that she did not have projects in which simulations were relevant. The VRE-D admitted that up to the time of the interviews, there had not been a simulation tool plugged into the Alfresco VRE, but mentioned that platforms such as HUBzero did have HPC built into it, which meant that one could run a simulation in it. The answers received indicated that simulations were not given such a high priority by the respondents, but it was possible to plug in a simulation tool. If the VRE for instance were to be transferred in the future to HUBzero, the simulation tool would be a built-in feature in the core function of the VRE.

(iii) Laboratory Experiments

Responses received from respondents in both case studies indicated that none of them had been doing laboratory experiments through the VRE. A-R1 mentioned that the way experiments were done at the time, had changed considerably. On the one hand, one still gets wet lab experiments that cannot be done on a computer, but on the other hand, one could do bioinformatics and programming experiments via a computer. She stressed that to use the latter type of experiments via a VRE might necessitate getting the rights from the people who developed the software. The VRE-D agreed that laboratory experiments could not be done within the Alfresco VRE, but mentioned that the data generated through the experiments could be uploaded manually. He again mentioned the HUBzero platform and stated that it would be possible to run experiments within HUBzero. The answers received from respondents showed that wet lab experiments would not be possible through the VRE, but the data flowing from that could be uploaded manually to the VRE, as was the practice at the time of the interviews. Experiments done via computerised instruments, as mentioned in 4.5.2.1 and 5.4, would be able to be done via a VRE, if these instruments are linked or plugged into the VRE. Some systems, such as HUBzero, have already been built into the platform.

(iv) Surveys

Survey tools such as Survey Monkey and Qualtrics were mentioned in 4.5.2.1 and 5.4 as examples of data capturing tools that could be plugged into a VRE. Eight of the respondents (A-R1, A-R3, B-R1, B-R2, B-R3, B-R4, B-VRE-M, and VRE-D) indicated that it would be possible to do surveys through the Alfresco platform, and six (A-R2, A-R4, A-R5, A-VRE-C, A-VRE-M, and B-L) indicated that it could not be done, or that they did not know, or that it was not applicable to them. A-R3 mentioned that in their field of research, they did not normally use surveys, which might also be why the majority of respondents in this case study indicated 'no' to this question. Her answer was very much in line with the nature of this case study, which focuses on natural science-oriented data, and laboratory/experimental methods, as mentioned in 7.2.1. On the other hand, the majority of respondents (B-R1, B-R2, B-R4, and B-VRE-M) in Case Study B indicated that they had been, or would be, using surveys in their research, which was very much in line with the nature of this case study, which focused on human-oriented data and used survey instruments as data collection method, as mentioned in 7.2.2. The VRE-D revealed that Alfresco didn't have a survey tool built into it, but that he had developed a survey platform, which he had then plugged into Alfresco (also see 7.2.2.4).

(v) Literature

Twelve of the respondents in both case studies (A-R1, A-R2, A-R3, A-R4, A-VRE-C, B-R1, B-R2, B-R3, B-R4, B-VRE-M, and B-L) revealed that they had been using the VRE to capture literature, and two (A-R5 and A-VRE-M) indicated that they had not been doing this. The A-VRE-C revealed that they had been uploading articles that they produced, but had also started a library with articles that they use as data. The B-VRE-M emphasised that literature searches and search strategies had been part and parcel of every research project they embarked on, and the B-L mentioned that in her capacity as librarian, she regularly did searches and uploads of literature onto the VRE. A-R5 indicated that she was not aware of the library the A-VRE-C mentioned. This could be a reflection of the fact that she had not attended the hands-on training session that the others in the group had undergone. During that session, the possibility of a library with articles was discussed. It could also indicate a lack of interaction or communication in the group, which might be because the social aspect of the group had not been developed yet. The A-VRE-M acknowledged that at the start of the VRE, there had been a discussion to set up an article/literature library where articles or other literature could be uploaded for future reference. He indicated that he had not placed this on the VRE yet, because every member of the VRE would use different keywords for different articles or literature, which makes it complicated to find something in the library. It would be quicker, according to him, to go to a database such as PubMed and search for the original, than to try and find an article on the library in the VRE.

The answers showed that there is a need for the VRE to capture literature, and the system makes provision for this through the documents store, as part of the core interface / software interface layer, as proposed in 5.4.

(vi) <u>Other</u>

The B-VRE-M suggested that a programme be added to the VRE, which would make it possible to do systematic reviews, because if the researchers had this programme available through the VRE, their data would also be captured on it. Up to the interview dates, the process of systematic review had been uploaded manually as a package. This component was not found in the literature and could be added to the conceptual VRE model as a pluggable component.

(b) Store / Backup data

Thirteen of the respondents (A-R1, A-R2, A-R3, A-R4, A-R5, A-VRE-C, A-VRE-M, B-R1, B-R2, B-R3, B-R4, B-VRE-M, VRE-D) indicated that they had been able to store or back up data using the VRE, and one respondent (B-L) indicated that she was unsure, which could be because of a lack of training on the system. A-R4, however, mentioned that he did not use the VRE exclusively to store or back up data. He stated that he also used other tools for storage and backup of his data. He indicated that he had been doing it to spread his risk across different tools. This signalled a motion of distrust in the VRE. The A-VRE-M, at the time of the interview, mentioned that the backing up and storage of data had been their main purpose for using the VRE. The VRE-D revealed that if the data were placed within their personal directories on the VRE, the data automatically were backed up and replicated. He also mentioned that there had been three servers, geographically separated, that replicated the data, and if the data were on the VRE, the data were automatically backed-up on all three. The document store in the core interface / software interface layer mentioned in 5.4, would be the ideal place to store or backup data in a VRE.

(c) Store different versions of data

Twelve of the respondents (A-R1, A-R2, A-R3, A-R4, A-VRE-C, A-VRE-M, B-R1, B-R2, B-R3, B-R4, B-VRE-M, B-L) indicated that they had been storing different versions of the data on the VRE and one (A-R5) indicated that she had not been doing that. This answer differed, however, from the answers received under ‘Upload a New Version’ under question 8, where it was mentioned that only three members in Case Study A and three members in Case Study B had used that component. This meant that the other members of these Case Studies must have been adding new versions of files themselves and were not using the ‘Upload a New Version’ component. This could perhaps be because of a lack of understanding of how the process works or could be due to a lack of training.

The VRE-D mentioned that members of the VREs could store different versions of a file, and then gave an example of a file that the members of Case Study A had been using regularly for refill-purposes of their liquid Nitrogen. This file, according to him, might already be in the region of version 50 of that document. The answer given by A-R5 showed that the reason she had not been using the versioning function was because of lack of knowledge about the system. This could be traced back to the lack of hands-on training that this respondent had received. The Alfresco system that the respondents had been using as a VRE has a versioning function built into it, which meant that the system automatically created different versions of a data file every time it was updated. Versioning was also one of the main reasons the Moodle platform was replaced by the Alfresco platform (see 7.2.1.8). The versioning function is something that would typically form part of a document store in the core interface of a VRE as mentioned in 5.4. It could also form part of a pluggable document management system, also mentioned in 5.4.

(d) Add Metadata

Two of the respondents (B-R2, B-R4) stated that they had been adding metadata and twelve (A-R1, A-R2, A-R3, A-R4, A-R5, A-VRE-C, A-VRE-M, B-R1, B-R3, B-VRE-M, B-L) mentioned that they had not been doing this. The metadata to which these two respondents referred, however, could not be the Dublin Core Metadata Template that was mentioned under question 8. These would most probably be categories, tags, and comments (see answers to question 8). A-R1 mentioned that she was not sure what metadata is. A-R2 indicated that she hadn’t been aware of metadata until the researcher of this study and the VRE-D had a meeting with them, where the issue of metadata was discussed. The issue of adding metadata, however, had not been followed up again. A-R2 indicated that they would need coaching on the appropriate metadata standard to use. A-R3 felt that she did not need to add metadata, as the volumes of data that she had been uploading were minimal. In other words, she knew where everything could be found. She did, however, mention that metadata might be useful when handling raw data, because one could sometimes lose track of which ‘repeat’ of a dataset worked or not. A-R4 stated that he used file-naming conventions to find files again, and therefore did not use metadata. A-R5 mentioned that she had been adding tags to the shared files, but stressed that these were very minimal. A-VRE-C admitted that one could add metadata, but that no one in her case study had really been doing it. The reason for the non-usage of metadata in Case Study A was summed up by the A-VRE-M. He mentioned that adding metadata would have a negative effect on the researcher’s time, as it would be too time-consuming. His solution to this would be to appoint someone fulltime to add the metadata on behalf of the researchers.

In Case Study B, B-R3 mentioned that she had not added metadata yet, because she had not really started using the VRE, but agreed that it would be a great help in finding a file again. The B-VRE-M stated that VRE group had reached the stage where they would need to learn how to do it. The B-L indicated that she was not aware that the researchers in Case Study B had been doing that. This could, however, just be a misunderstanding from her side on what constitutes metadata, because the B-R2 indicated that the B-L assisted her in adding metadata to the files. The B-R4 indicated that they added metadata to data that came from the 3D Motion Capture. These metadata helped them to identify data without mentioning research respondents' names. The VRE-D confirmed that researchers could add metadata to a file, but also mentioned that the VRE system (Alfresco) pulled technical metadata from the system, e.g. an image would include the resolution, the size, the file type, and even geographical location (e.g. GPS coordinates), etc. The researcher of this study then asked the VRE-D how one could ensure that respondents add metadata. He mentioned that he could set up the system in such a way that it would force the researchers to complete certain metadata fields upon ingestion of certain file types. He was nevertheless concerned that it would make the system too time consuming, and would cause researchers not to use the VRE. The low uptake of adding metadata, which was also mentioned under question 8, is concerning, because metadata are important for the discovery (retrievability) and for the understanding of the context and provenance of data (see 4.5.2.7), as well as the re-use of data, which is one of the stages in the research data lifecycle (see Figure 4.7 and 4.5.2.6).

The Alfresco system made provision for adding bibliographic metadata, but also automatically added technical metadata to the data files. This would typically form part of the document store in the core interface or could be part of the pluggable document management system, as indicated in 5.4. To obtain compliance by researchers to add metadata, adding of metadata could typically be included in the policy components layer as part of the ground rules or an RDM plan (see 5.4 and Figure 5.2e).

(e) Process data

In 7.2.1.5, the members of Case Study A expressed a wish to be able to access their processing and analysis programmes within the VRE. At the time of the interviews, however, all the respondents from both case studies indicated that they could not do processing of data through the VRE platform (Alfresco). A-R2 and B-R1 notwithstanding, mentioned that they were of the opinion that there might be a processing type of programme that could be plugged into the system. This was confirmed by the VRE-D, who indicated that Alfresco did not currently have this capability, although it could be customised quite extensively. He then mentioned as an example the HUBzero VRE platform, which in its vanilla setup have the capability to process data. This links to the data processing stage of the data lifecycle as mentioned in 4.5.2.2.

(f) Analyse data

All the respondents from both case studies indicated that they could not, at the time of the interviews, do analysis of data through the VRE platform (Alfresco). (This was a wish that was expressed in 7.2.1.5). A-R1 mentioned that this had been one of the functionalities they had asked for from the start, but that the amount of computing power needed might cause the programme to crash. The B-VRE-M indicated that with one of the clinical projects, they had reached the stage where they would need to use analysis software, and that it would be preferable if this could be built into (plugged into) the VRE platform. The VRE-D confirmed that it was not currently possible within the Alfresco VRE platform, but was already available within the HUBzero platform. Data analysis tools were also mentioned in 5.4 and 4.5.2.3 as RDM tools that could be added to a VRE.

(g) Visualise data

Although visualisations of data were not an Alfresco functionality, the purpose with this question was to find out if a visualisation tool had been plugged into the system.

All the respondents from both case studies indicated that, at the time of the interviews, they could not do visualisations of data through the VRE platform (Alfresco). A-R2 stated that she did not do visualisations of her data. A-R3, A-R4 and A-VRE-C mentioned that they had been generating visualisations of their data, but had not been doing this through the VRE platform, because the Alfresco VRE did not have the functionality yet. A-R4 indicated that it would be valuable to have a visualisation tool available within the VRE, but expressed that it might be very specific to equipment attached to an instrument. A-R4 also suggested adding a tool that would be able to generate visualisations of statistics (management information), for example a graph in a monthly report that kept track of how many times equipment had been used per day or per month, or which person had used it the most, etc. The B-VRE-M indicated that the projects within the case study had reached the stage where they would need to be able to generate visualisations within the VRE. The VRE-D admitted that it could not be done in the Alfresco platform, but was also not certain if one could do this within HUBzero. He was of the opinion, however, that it would be possible to do customisations either in Alfresco or HUBzero, which would enable one to do it. Visualisation tools were mentioned in 4.5.2.3 and 5.4 as possible RDM components in a VRE.

(h) Share data, with peers or with supervisor (Workflow)

Nine of the respondents (A-R1, A-R3, A-R5, A-VRE-C, B-R1, B-R2, B-R4, B-VRE-M, B-L) mentioned that they had been sharing data and four (A-R2, A-R4, A-VRE-M, B-R3) stated that they had not been doing this. A-R2 indicated that the A-VRE-M, as well as a post-doctoral fellow that is collaborating on this project, had access to her files, but they did not use the workflow function, probably because this way of working (using shared folders and files) had been sufficient for them. A-R4 mentioned that he had shared files via e-mail and even Mendeley. This corresponded with his earlier remarks that the system is not as user-friendly as he would have liked it to be, and that he preferred spreading his files and data across a number of tools. The VRE- D confirmed that the functionality had been available via Alfresco for those respondents who wanted to use it. Through observation of the case studies, the researcher of this study also found that respondents within Case Study B had been using this component quite extensively between the B-VRE-M and the rest of the respondents. The Alfresco platform has a built-in workflow system as part of its core interface, but one could also add a data workflow system to the VRE, as mentioned in 5.4.

(i) Publishing of data in a repository? If not, are you publishing your data elsewhere?

In 4.5.2.3 and 5.4, publishing of data in a data repository was mentioned as a component that could be added to a VRE. This question was asked to determine if the respondents are publishing their data into a repository through the VRE. All of the respondents in Case Study A and B indicated that they had not yet published data in a repository. A-R3 indicated that she had not been at the publishing stage of her research yet, whereas A-R4 stated that he had been aware of the movement to publish data, and how important it was when they do an experiment, to already have an idea how they will structure the output of that. He further expressed a need that the Library come and advise them on data publishing and repositories. The A-VRE-C indicated that their group had not really been at the stage where they had the data that should be published on a repository, ready. The A-VRE-M confirmed that the members of the group had not been publishing their data. They had only been publishing the outputs derived through the analysis of their data. B-R2 commented that it was possible to publish data in a repository, but that she had been busy preparing the first article that would be published elsewhere.

B-R3 mentioned that she had not given this any thought, but stated that it would be good to share it. B-R4 emphasised the need for the publishing of data that accompanied articles. This, according to her, had been fuelled by the requirements coming from publishers, for the publishing of data in repositories. She also mentioned that there had been no discipline-specific repository for her research field available. She indicated further that some of the publishers had repositories available, but these had been very expensive to use. She also expressed her need for a data repository for the University. The B-VRE-M mentioned that their VRE group were very close to the stage where they would publish their data, which had been on Alfresco, into a repository. She then revealed that in one of the projects they did publish the pilot study, with a paragraph indicating that they would make the data available when the full study was published. The VRE-D confirmed that at the time of the interview, the Alfresco system did not make provision for publishing in a repository. The system, according to him, only allowed publishing to social media such as Facebook or LinkedIn, and would have to be customised to make provision for publishing to repositories. He indicated that DSpace and Fedora, for example, use web APIs, which made it possible to get access to the code that would allow another system such as Alfresco to post data into the repository system. The code in Alfresco, according to him, could also be customised to do that.

(j) Preserve data for long-term

In 4.5.2.4, the researcher of this study discussed the preservation stage of the research data lifecycle. One of the components of this stage is long-term preservation. Although data preservation was not part of the functionality of the Alfresco system, the aim with this question was to gauge whether the respondents had been preserving their data for the long-term by using a preservation system that had been plugged into the VRE platform. The answers showed seven of the respondents from both case studies (A-R3, A-R5, A-VRE-C, A-VRE-M, B-R1, B-R2, B-R4) were under the impression that the data that they had been uploading onto Alfresco had automatically been preserved for the long-term. This revealed a misunderstanding of the concept 'data preservation'. The data that they had been uploading on Alfresco did not meet the requirements for data preservation, namely providing enough representation information, context, metadata, fixity, etc. to the data so that anyone other than the original data creator could find, use and interpret the data (Choudhury, 2014: 125, cited in 4.5.2.4). The rest of the respondents (A-R1, A-R2, A-R4, B-R3, B-L, B-VRE-M, and VRE-D) indicated that they had not been doing this. A-R1 mentioned that she had not used it yet, but stated that it was something that she would use in future. She also mentioned that her supervisor (A-VRE-M) had the responsibility to see that their data were kept for at least 20 years. A-R4 showed through his answer that he also did not understand the concept of data preservation. He had been under the impression that if he saved various versions of a file on different platforms or devices, (e.g. hard drive of a computer, or cloud), then he preserved his files. The answer received from B-R3 showed that she was not certain what the concept of data preservation meant, but that she was aware that they had to keep data for 15 years, and that she planned to do that once her study proceeded. The B-VRE-M indicated that their group would be preserving their data. The answer given by the B-L revealed confusion about the terms back-up and preservation. The researcher of this study then explained the difference between the two concepts to her. The B-L subsequently mentioned the possibility of using a microfiche for long-term preservation, which is an old format that was used in the past to capture and preserve information. This revealed a lack of knowledge about current forms of electronic preservation methods and formats. The VRE-D revealed that Alfresco did not by default do automated data preservation. He reiterated again that at the time of the interview, he was testing it manually. When a project on Alfresco reached completion, he bagged and tagged it with BagIt, manually. The manual method would, however, need skilled human resources that had not been available. He mentioned that an automated process would be preferable, and that repositories that are part of VREs would probably in the future have BagIt built into them (as a core function). This would make it possible to be able to specify dates, times and locations where data would be preserved. The alternative, although not specifically mentioned by the VRE-D, would be to plug-in a data preservation component into the VRE (seeing that Alfresco has an open API), which is in line with data preservation mentioned as an RDM component in 5.4. The answers from the majority of respondents further showed that some training with regards to long-term preservation would be needed.

(16) Is the use of the data restricted by the following?

(a) Confidentiality / ethical reasons

(b) Law

(c) Proprietary / commercial interests

(d) Creative Commons License

The aim with this question, as mentioned in 6.3, was to determine the openness of the data that are managed. This would have an impact on the sharing of data, which was mentioned in the 'giving access to data stage' of the research data lifecycle in 4.5.2.5.

(a) Confidentiality / ethical reasons

All of the respondents in both case studies indicated that their data had been restricted for confidentiality and ethical reasons. A-R1 indicated that some of her data were not accessible to people outside the VRE (in the public domain), because of confidentiality reasons, but that all her data had been open to the members of the VRE group. A-R2 also stated that her data had been restricted because of ethical and confidentiality reasons. The method she had used was to assign alphanumeric codes to the persons in her study in order to de-identify them. This was also stated in her application for ethical approval. She expressed, however, her concerns whether this had been sufficient to safeguard confidentiality, because of developments in web technology that could potentially trace data back to an individual. The assignment of alphanumeric codes to personal data corresponds to the discussions in 4.5.2.5 on unlinked anonymised data that include no information that could sensibly be utilised by others to identify persons. It is also in line with the discussion on a de-identified dataset in 4.5.2.5, where the data can be key-coded, encrypted, or pseudonymised to remove personal information. A-R5 mentioned that her data had not been restricted because of ethical reasons, but definitely because of confidentiality reasons, for example if one worked on something and did not want to share that completely, in other words, just share a limited dataset (see 4.5.2.5), or sometimes first publish an article, and then discuss the data with peers that could understand the context. The A-VRE-M stressed that there should be a balance between de-identification and traceability, because under certain circumstances, one has to be able to trace an individual, for example, if there is susceptibility to a certain disease. His answer is in line with the discussion in 4.5.2.5 on the reversible type of process that can be followed when setting up a de-identified dataset. Although the information that can identify a person is key-coded, encrypted, or pseudonymised, this process can be reversed under certain circumstances, so that a research group can do research on or trace the spread of a specific disease, etc., for example.

B-VRE-M indicated that the person(s) that do(es) the de-identification, be varied between institutions. The answer given by B-R3 corresponded with the answer given by A-R2, namely the assignment of code numbers to respondents so that individuals cannot be identified. Information on individuals should, according to her, just be limited to the research project and not shared with others, which is more in line with the limited and protected dataset in 4.5.2.5. B-R4 indicated that what they had uploaded onto a system such as Alfresco, were strictly controlled by the Ethics Committee of their Faculty. She then also shared that Alfresco had a confidentiality clause that enabled one to store files confidentially. This meant that only she and her supervisor could see things that are very confidential. If she then wanted to publish the data, she had to ensure that the data were anonymised, a process mentioned in 4.5.2.2 in the 'processing of data stage' as well as in 4.5.2.5 in the 'sharing of data stage' of the research data lifecycle. The B-VRE-M stated that data that were published (in other words, in the public domain) were not restricted by confidentiality.

(b) Law

All the respondents in Case Study A mentioned that there were legal restrictions on their data, except A-R5. A-R1 mentioned that there were some things by law that they were not allowed to disclose about their respondents. A-R2 confirmed this and indicated specifically that they were not allowed to make personal information of respondents available. A-R4 mentioned that the use of facial material had legal restrictions, and that personal information was also protected by the POPI Act. The A-VRE-C mentioned that there were legal restrictions if they were to develop certain treatments, or drugs. The A-VRE-M stated that there might be intellectual property issues that could have legal implications. In Case Study B, only B-R1 stated that there were some legal restrictions to her data, but the rest of the respondents indicated that they had no legal restrictions to their data. These restrictions could, however, be overcome by anonymising the data, a process mentioned in 4.5.2.2 in the processing of data stage of the research data lifecycle.

(c) Proprietary / commercial interests

In 4.5.2.5, it was mentioned that the sharing of data could be viewed as a valuable "component of the scientific process" and that the sharing of data affords "opportunities for other researchers to review, confirm or challenge research findings" (Institute of Education Sciences, n.d.). In some cases, data can, however, be restricted for proprietary or commercial reasons. This can, according to 4.5.2.1, be given in the metadata to a data set, where the access conditions and terms of use of a data set can be described. Restrictions can also be set by using a copyright license, for example a Creative Commons License.

In Case Study A, all the respondents mentioned that their data were restricted because of proprietary or commercial interests, to some or other degree. A-R1 indicated that some in the group were waiting for patents to come from the products that they had been working on. She also mentioned that she had been looking into commercialising something out of the project that she was working on. Her data would only be made available once that had happened. A-R2 and A-R3 confirmed the idea of patents that were being registered for some of the products that had been developed through their projects. A-R4 stressed that this did not apply to everything they did, but that there were certainly restrictions to some of their data, because of proprietary or commercial interests. The A-VRE-C confirmed that there were a number of projects in the group where the data had been restricted because of proprietary or commercial reasons.

The answers of respondents in Case Study B revealed that none of the respondents' data had been restricted because of proprietary or commercial interests.

(d) Creative Commons License

A Creative Commons License is a copyright license that affords people a simple, "standardized way to grant copyright permissions to their creative work" - in this case, their data (Creative Commons, n.d.). Each license will determine what could be done with the data (see also 4.5.2.5).

The answers received showed that none of the respondents in either of the case studies had published their data or results yet, and none of them had even considered the possibility of publishing their data. This could be a because of a lack of knowledge about the process of data publishing. It also meant that at the time of the interviews, no one had yet added a Creative Commons License, which could have restricted the use of their data.

(17) What methods / tools do you use to analyse the data?

The methods and tools used by respondents in each case study were listed in the first columns of Tables 7.3 and 7.4 and then matched to the respondents that used these, in the second column.

The purpose with this question, as indicated in 6.3, was to determine what tools the respondents were using to analyse their data.

The results from both case studies showed that the two groups had been using different tools that were more characteristic of the nature of the disciplinary areas of each group.

For example, respondents from Case Study A had been using instrument-based software (e.g. Kaluza), Galaxy, Statistical Analysis Packages (e.g. Statistica, SPSS, and R), whereas respondents in Case Study B had been using interviews, questionnaires, surveys, analytical schemas or algorithms, literature searches, AGREE II tool, systematic reviews (e.g. Eppi Reviewer and Revman), Video Nystagmography, a machine that tests the inner ear, a machine that measures speed, Qualysis Track Manager, and qualitative data-analysis programs (e.g. Atlas.ti). Tools that respondents in both case studies used included Microsoft Office tools (e.g. MS Excel, MS Word), and Statistical Analysis Packages, (e.g., SPSS).

The results from Case Study A further showed that most of the respondents had been using more than one tool to analyse their data. For example, A-R1 had been using MS Excel to do quick trimming and cleaning of her data before she used R (a statistical analysis package), to process and analyse her data. A-R1 also used instrument-based software, such as Kaluza, as well as CLC Genomics Workbench, and A-R2 used MS Word, MS Excel, CLC Genomics Workbench, Galaxy, and Variance Effect Predictor Tool.

Table 7.3: Methods And Tools Used To Analyse Data In Case Study A

Method / Tool	Respondent
MS Word	A-R2 A-VRE-C
MS Excel	A-R1 used MS Excel to do quick trimming and cleaning of her data before she analysed her data with R. A-R2 A-R4 A-R5 A-VRE-C
Flow cytometry analysis	A-R1 A-R3 A-R5 A-VRE-C
Statistical analysis	A-R1 analysed all her data in R. A-R4 used Statistica, and SPSS, and R for data analysis. A-R5 used specific statistical analysis packages for statistical analysis.

	A-VRE-C used a number of statistical programmes, but especially R.
Computational Analysis	A-R1 A-R2 A-R3 A-R4 A-R5 A-VRE-C
Data Cleansing	A-R1 used MS Excel to clean her data A-R5 used MS Excel to clean her data
Generation of custom workflows	A-R2
Variant detection and multiple filtering	A-R1 A-R2 A-R3 A-R4 A-R5 A-VRE-C
Instrument-based software, e.g. Kaluza	A-R1 A-R3 used Kaluza for flow cytometry data. A-R5 used instrument-based software such as Kaluza to process her data. The data were then exported to MS Excel where the data were cleaned up. She subsequently did statistical analysis on certain statistical analysis packages. A-VRE-C used Kaluza for analysis of flow cytometry data.
CLC Genomics Workbench	A-R2
Galaxy	A-R2 A-R3
Variance Effect Predictor Tool	A-R2
Statistical Analysis Packages, e.g. Statistica, SPSS, R	A-R1 processed and analysed all her data in R. A-R4 used Statistica, and SPSS, and R. A-R5 used specific statistical analysis packages. A-VRE-C used a number of statistical programmes, but especially R.

Table 7.4: Methods And Tools Used To Analyse Data In Case Study B

Method / Tool	Respondent
Interviews	B-R1
Questionnaire / Survey	B-R1
Analytical schema or algorithm	B-R1
Literature Search	B-R1
AGREE II tool	B-R1 used this tool to look at the quality of the research.
Objective Measurements to analyse the quality of a study	B-R1

Systematic reviews, e.g. Eppi Reviewer and Revman	B-R2 planned to use Eppi Reviewer, but indicated that it was very expensive. B-VRE-M was using Revman.
Video Nystagmography	B-R3 measured different eye movements through Video Nystagmography.
Machine that tests the inner ear	B-R3 used a machine that tests one's inner ear, one's reflexes down to one's neck. She indicated that this machine would be used together with the Video Nystagmography, to determine what the prevalence of visual vestibular disorders in stroke patients are. The software of these two instruments test and give quantified measurements, and comes out with graphs, tables, and video recordings.
MS Excel	B-R4, B-VRE-M, B-L
Statistical Analysis Packages, e.g. SPSS	B-VRE-M
Measure speed tool	B-R4
Qualysis Track Manager	B-VRE-M
Qualitative data-analysis programs, e.g. Atlas.ti	B-VRE-M

In Case Study B, some of the respondents had been using more than one tool to analyse their data, for example B-R1 indicated that she had been using interviews, questionnaires, surveys, analytical schema or algorithm, literature searches, the AGREE II tool to analyse her data and also objective measurements to analyse the quality of her study. Others indicated that they had been using only one tool, for example B-R2, which indicated that she had been doing systematic reviews by utilising Eppi Reviewer and Revman. None of these data analysis tools were found to be integrated in the VRE, but all of these, depending on their software licensing, could potentially be added to the VRE as RDM tools (see 5.4. and Figure 22 b).

(18) Do you use visualisations to analyse your data? If so, give examples.

The respondents in each case study were listed in the first columns of Tables 7.5 and 7.6 and then matched to the visualisation tools or procedures in the last columns. In the second and third columns, it was indicated whether the respondent used the tools or not.

Table 7.5: Visualisation Tools Or Procedures Used By Respondents In Case Study A

Respondent	Yes	No	Tool / Procedure
A-R1	✓		R
A-R2	✓		• Graphical Interfaces, which allows one to assess the quality of a sequence. • GL Pictures – by studying a picture, one makes a decision whether or not an experiment was successful.
A-R3	✓		Microscopy images.
A-R4	✓		Flow Diagrams, Images etc. to analyse one's data.
A-R5	✓		In Flow Cytometry, there is a large amount of visual input in the beginning, using specialised software. The results from these are then exported in the form of numbers to an Excel Spreadsheet.
A-VRE-C		✓	"No, I am not there yet".

Table 7.6: Visualisation Tools Or Procedures Used By Respondents In Case Study B

Respondent	Yes	No	Tool / Procedure
B-R1	✓		Algorithms are first plotted so as to form a picture, before an analysis / processing can be made. No further visualisations are done.
B-R2		✓	
B-R3	✓		As part of the visual tracking that will be done, the videos of the test subjects' eye movements will have to be studied and analysed. Graphs and images will also be created to assist in the analysis of the data.
B-R4		✓	
B-VRE-M	✓		Pictures or conceptual frameworks or different forms of graphic displays can be used in quantitative research projects. Pictures can also be used to explain data collection procedures / environments.
B-L	N / A	N / A	

As mentioned in 6.3, this question was asked to determine if any of the researchers used data visualisations to analyse their data. The answers received from respondents in Case Study A showed that five of the respondents (A-R1, A-R2, A-R3, A-R4, A-R5) used visualisations to analyse their data, and one (A-VRE-C) was not using this yet. Answers received from respondents in Case Study B showed that three of the respondents (B-R1, B-R3, B-VRE-M) had been using visualisations to analyse their data, while two (B-R2 and B-R4) indicated that they had not been using visualisations, and one (B-L) indicated that the question had not been applicable to her. Data visualisation was shown to be part of the 'analysing data stage' of the research data lifecycle in 4.5.2.3, and data visualization tools were also mentioned as possible RDM tools in 5.4. The responses received from respondents showed that the majority of members of Case Studies A and B used a variety of data visualisation tools, and that this would be a component that should be included as an RDM tool in a VRE conceptual model.

(19) Are RDM-related issues formalised within your VRE (for example, strategic action plans)? If yes, to what extent do you adhere to the policy?

All the respondents in both case studies indicated that there had been no formal policy in place with regards to RDM. It would seem that everyone had a general idea what to do. A-R3 indicated that there was a need for such a policy, because it would compel people in the group to use the Liquid Nitrogen database (folder). It would also promote the sharing of articles among the members of the group. A-R5 shared that there had been an indirect agreement in the group on what to do when someone leaves the group. For example, everything one had done should by then be on the VRE, and that should get tracked and monitored. The A-VRE-C shared that everyone had a general idea of how to store or backup their data, but they did not have a formal policy in place. She suggested that a formal policy with regards to metadata might be needed. The A-VRE-M also mentioned that they did not have something formal in place. He shared that there existed a vision and strategy for the research group, but he regarded, however, the VRE as just a point along the line of developing the research project. He did not see a formal policy as conducive at the time of the interview, as things were very fluid and they needed the fluidity to operate. From the answers received from respondents in Case Study A, it would be possible to conclude that there was a mutually agreed project plan and collaborative agreement on how data should be uploaded onto, and archived in the VRE, and shared through the VRE. This is in line with one of the policy components mentioned in 3.5.7.6 and in Figure 5.2e.

In Case Study B, B-R1, B-R2, and B-R3 indicated that it was outlined in their protocols what to do. These directions in their protocols were, however, subject to changes as they went along. B-R3 mentioned that it was specifically stated in her protocol that they should upload their data onto Alfresco and that she would preserve it there for 15 years. She mentioned that she had also been working together with a researcher from another discipline area, and had not been sure what the policy was with regards to her co-researcher’s data. B-R4 stated that the tools and type of data that they had collected differed quite a bit, which meant that everyone did their own thing. They just had to make sure that their supervisor (B-VRE-M) knew what they were doing. The B-VRE-M mentioned that they had actually been planning to put something formal in place, but that it had been a bit premature. It needed to be aligned with their departmental strategic plan. The B-L emphasized that she had not been part of the initial formation and training of the group, and therefore was not aware of a policy. She nevertheless thought that there should be something formal in place, for example a protocol.

The majority of respondents mentioned protocols as formal documents that guide their actions within the VRE, and although this was not mentioned in the literature study as a possible policy component (see 3.5.7.6 and Figure 5.2e), it is something that should be added to the VRE model (see Figure 5.2a). These protocols, however, did not include DMPs. Moreover, none of the respondents mentioned the creation of DMPs, that could have given an overview / parameter of the research that he/she planned to undertake, and also what he/she planned to do with his/her data. The issue of DMPs was discussed, however, at a higher level during a meeting on 11 April 2013 between members of the Library Services, the Deputy Dean Research of the Faculty to which the research group belong, the Chair of the Ethics Committee of this Faculty and the A-VRE-M (see 7.2.1.1). It would seem, however, that it was never followed up and applied in these two case studies.

Data management planning was mentioned in 5.2 as an action that should take place during the creating data stage and re-using data stage of the research data lifecycle.

The answers received from respondents in both case studies revealed furthermore that there was no RDM policy in place. The reason for this was that the VRE Managers were of the opinion that, at the time of the interviews, the VRE groups were not ready for this. The A-VRE-M felt that they needed fluidity, and the policy would inhibit development of the VRE, and the B-VRE-M felt that it was a bit too premature to put a policy in place. This could perhaps also be why DMPs were not mentioned by any of the interviewees. Another possibility could be due to a lack of knowledge on what a DMP is. To ensure the usage of DMPs, it should preferably be linked / integrated with the ethical process of the Faculty and University.

(20) Describe what other functionalities you can use in the VRE. Are you utilising these? If so, which functionalities? If not, why not?

As mentioned in 6.3, the intention with this question was to determine the awareness of the respondents about the various functionalities / tools of the VRE, as well as to gauge the uptake of these functionalities / tools.

The respondents could not really think of any other functionalities than had been discussed in the interviews. A-R3 mentioned that she had not used the VRE long enough to be able to mention all the functionalities. The A-VRE-C felt that almost everything had been discussed during the interview, and could not mention anything more. The A-VRE-M could not give feedback on the functionalities because he had not been using the VRE much. B-R1 felt that she had not explored the VRE system further and could therefore not give feedback. B-R3 mentioned that she had not used the system long enough to be able to give feedback. B-R4 stated that the system included nearly everything, and then added that the system had a valuable functionality of sending an e-mail alert the moment something happens on the system. The B-VRE-M could not think of something more, and the B-L mentioned that the product was still very unfamiliar to her, and she would need to experiment with it more in order to be able to provide feedback.

(21) What would you say are the objectives of the VRE within which the research project is managed?

(a) Immediate objectives

(i) <u>Use the VRE for storage / back-up of data</u>

The majority of the respondents (A-R1, A-R3, A-R4, A-R5, A-VRE-C, A-VRE-M, B-R1, B-R2, B-VRE-M, and VRE-D) mentioned that the immediate objectives of the VRE were storage or backup of data. This corresponded with 5.3, where it was mentioned that Carusi and Reimer's (2010: 18-19) Virtual Research Environment Collaborative Landscape Study viewed a VRE as an easy-to-use platform where researchers can secure the short-term storage of their data. This was also confirmed by Neuroth, Lohmeyer and Smith (2011: 225), as cited in 5.3. According to them, a VRE will provide researchers with a safe place where they can save their data directly.

(ii) <u>Use the VRE for retrieval and access to data</u>

Some respondents (A-R1, A-R2, A-R5, A-VRE-C, B-R1, B-R-2, B-R3, B-VRE-M) indicated the recovery and retrieval of, and access to their data, as some of the immediate objectives, which were in line with Carusi and Reimer's (2010: 13), view of a VRE as "providing access to 'data', tools and services through a technological framework," as mentioned in 5.3. Some of the respondents (B-R2, B-R3 and B-L) even mentioned the idea of accessing the VRE remotely from another geographical area. This corresponded with 5.3, where it was mentioned that Carusi and Reimer (2010: 22) and Anderson, Dunn and Hughes (2005: 516) saw VREs as bringing together geographically dispersed researchers.

(iii) <u>Use the VRE as a common centralised space / framework</u>

The idea of having a common, centralised space, system or location was also broached by some respondents (A-R2, A-R3, A-R5, VRE-D, B-R3, and B-L). This was in line with the researcher of this study's definition of a VRE in 2.2.8.1, where he defined a VRE as a "common, flexible, technological and collaborative framework."

(iv) Use the VRE for data-sharing

Data-sharing was another objective that was mentioned by some of the respondents (A-R3, A-R5, B-R1, B-R3, and B-VRE-M). Carusi and Reimer (2010: 19), as cited in 5.3, mentioned this sharing of data aspect of VREs, while Filetti and Gnauck (2011: 237) emphasized that data sharing is a key element in a VRE.

(v) Track the progress of research students, through the VRE

A number of respondents (A-R2, B-R1, B-VRE-M, and B-L) identified the functionality that the VRE provided to the supervisor, to track their student researchers' progress, as an immediate objective. The A-VRE-M, however, was not fully engaged in the VRE, which meant that he did not mention that he was tracking the progress of his students.

(vi) Provide access to protocols of experiments, via the VRE

A-R1 mentioned the idea of making the protocols of different experiments available to other members in the group, via the VRE.

(vii) Preserve data through the VRE

Another objective that was mentioned was the preservation of data, which corresponded with Carusi and Reimer's (2010: 18-19) Virtual Research Environment Collaborative Landscape Study, mentioned in 5.3, which showed that integrating architecture for data management within a VRE, can address the issue of preservation of research data. The A-VRE-C and B-R3, however, saw the Alfresco VRE as a place where data could be preserved for long-term. This indicated a misunderstanding of what long-term data preservation entails, as the Alfresco system did not have a long-term preservation function built into it.

(viii) Work collaboratively through the VRE

A-R3 added the aspect of working together collaboratively on a document in Alfresco, as another objective. This collaborative nature of VREs was mentioned in 5.3 as providing researchers with the possibility to "share data and collaborate" in collecting, manipulating, analysing and interpreting data (JISC, 2006; Carusi and Reimer, 2010: 20).

(ix) Use the VRE to provide a secure space for one's data

The VRE-D touched on the aspect of securing data and the B-VRE-M mentioned that the VRE could prevent data-loss. This was in line with 5.3, where it was mentioned that Carusi and Reimer (2010: 18-19) and Neuroth, Lohmeyer and Smith (2011: 225) emphasized the VRE characteristic of uploading data in a safe and secure place.

(x) Replace paper-based processes with online processes provided through the VRE

B-R4 stated that for her, the immediate objective was to replace the paper-based system that they had used before, with an online system.

(xi) Use the VRE as a communication platform

The B-L raised the issue of communication, and indicated that the VRE provided a much easier way of communicating than e-mail. The system kept track of all communications in one place, and was much more organised than e-mail. This corresponded with one of the characteristics of a VRE that was mentioned in 2.2.8.3, namely that "a VRE system should be able to act as communication platform" (Yang and Allan, 2006a: 453; Wilson, et al., 2007: 290).

(b) Is there a time limit for the VRE?

The majority of the respondents (A-R1, A-R2, A-R3, A-R4, A-VRE-C, B-R2, B-R3, B-R4, and B-VRE-M) were of the opinion that there was no time restriction on the VRE, and that it would continue to be used. A-R1 felt that the group would keep on using the VRE for storage of data, but expressed her concern that the current VRE platform, Alfresco, would not be able to handle the uploading, downloading and data mining of large data - an area into which the group had been expanding. She indicated that they had discussed this with the VRE-D and that he had mentioned the possibility of using virtual machines to handle big data. A-R2 indicated that she had not been aware of any time restrictions on the VRE, and was of the opinion that it potentially could be something that everyone would eventually use. Two of the respondents (B-R1 and A-R5) were of the opinion that the VRE might come to an end when they finish their projects. A-R5 stated her doubts about the current VRE platform. At the time of the interview she was not sure if people would easily buy into the system. Her answer reflected the fact that she had not received the full training that the other respondents had received, which made the system difficult for her to use. It could also be an indication that the system might not have been as user-friendly as one would have wanted it to be. The A-VRE-C revealed that the VRE system was continually developing, and stated that she hoped that the system would at some or other stage reach a level where it would be stable. The A-VRE-M saw the VRE as the start of a much bigger development in the research process.

B-R3 indicated that even when they finish with their individual projects, they would still want to have access to it for clinical practice. B-R4 was of the opinion that the VRE would just keep on growing. The B-VRE-M also mentioned that she was of the opinion that there was no time limit to the VRE, but indicated that the individual projects that had been running on it, would still run for another two years at least. The B-L stated that she was not sure what would happen when the researchers in the group graduated. The VRE-D was of the opinion that the methods in VREs, and how they operate, were actually quite logic, and that what is now called VREs would in a few years’ time be quite standard features found in an organization.

(c) Objectives beyond the project period

The answers received from the respondents proved to be very insightful. The respondents identified the following objectives:

(i) Provide data processing through the VRE

A number of respondents (A-R1, A-R3, A-VRE-C, A-VRE-M, and VRE-D) indicated that they would want to see data processing included in the VRE. This was in line with the 'data processing stage' of the research data lifecycle as mentioned in 4.5.2.2 as well as 5.3, where it was mentioned that VREs could be used for data processing (Carusi and Reimer, 2010: 19).

(ii) Analyse data through the VRE

A-R1, A-R3, A-VRE-C, A-VRE-M, and VRE-D also indicated that they would like to see the analysis of data (mentioned by A-R3, and VRE-D), as well as sequencing and simulations (mentioned by VRE-D), to be included. Analysing data is mentioned as one of the stages of the research data lifecycle in 4.5.2.3, and is also mentioned in 5.3 as a process that can be done through a VRE (Carusi and Reimer, 2010: 19; Filetti and Gnauck, 2011: 238). Martinez-Uribe and MacDonald (2009: 311), as mentioned in 5.3, found that research data generated inter alia through "models / simulations, are intrinsically linked with data collection methodologies and instrumentation," and that a VRE is the ideal place to position it.

(iii) Create capacity in the VRE to enable it to handle big data sets

A-R1 wanted the ability of the VRE platform to handle big data sets (see discussion on big data in 4.6).

(iv) Add processing and computing power to the VRE

A-R3 and A-VRE-C mentioned that the VRE platform would need more processing power and more storage space. This was subsequently confirmed by the A-VRE-M, who mentioned that they would need more computing power. This corresponded with 5.3, where it was stated that VREs provide easy access to computational resources and collaborators, resulting in "faster research results and novel research directions" (Carusi and Reimer, 2010: 5; Pham et al., 2005: 16).

(v) Use the VRE for multidisciplinary research

A-R2, A-R5, B-R5 and B-VRE-M touched on the issue of using the VRE for multidisciplinary research, where researchers across different research groups share information and expertise and interact (A-R5, B-R2, B-R4, and VRE-D). This would include things such as using secondary data created by other researchers. This issue of multidisciplinary research confirms one of the characteristics of a VRE that was mentioned in in 2.2.8.3, namely that VREs enable inter-disciplinarity by bringing data and approaches from different disciplines together to "create new research findings" (Carusi and Reimer, 2010: 23, Fraser 2005).

(vi) Establish a formal structure for RDM at the University

A-R4 expressed the need for a more formal structure for data management at the University. He touched on things such as the infrastructure that should be readily accessible [integrated] in terms of their daily routines. He also mentioned the need for a University policy, which should clearly stipulate that the responsibility lies with the researcher, to make sure that his/her data are safe and secure. Such a policy, according to him, should specify that the data actually belong to the University, and stipulate that the University would provide the necessary facilities and funding. These comments on the need of a data management policy showed a lack of knowledge about the University policy that has been in place since 2007 (see 4.8.1). His comments also showed that the policy of 2007 was lacking and that a new policy would be needed, as mentioned in 4.8.7.

(vii) Establish a VRE system with an interface similar to Facebook

A-R5 declared the need for a VRE platform or system that would have an interface similar to Facebook. Such a system, according to her, should be much more intuitive. This corresponds with one of the characteristics mentioned in 2.2.8.3, namely that researchers would expect Web 2.0 (e.g. Facebook) and semantic Web technologies in a VRE (Yang and Allan, 2010: 68).

(viii) Establish a high performance-computing set-up

The A-VRE-M elaborated on his future objectives for his group (Case Study A). He stated that they had been generating huge numbers of data and had received funding to acquire some very sophisticated and expensive equipment. Part of this was a number of desktop computers that had good storage capacity and computing power. These desktop computers met their need for computing power and storage capacity temporarily. He then elaborated on the future, and mentioned that they had negotiated with Dell for the purchase of a HPC setup. This would be installed in the near future. This setup would have very high processing capacity, but he emphasized that solving the issue of storage / backup of such big amounts of data would be a challenge. What would need to be negotiated is whether such a storage facility would be tangible in a server room, or whether it would be in the cloud. He added that it would not just be about the storage of the data, but also about the management of that data, in other words, making it accessible, which would require careful thinking and planning, and might even require a full-time staff member to administrate it for them. Things that would need to be addressed in such a system would be multiple levels of security, managed access, a data access policy, etc. The A-VRE-M also saw the current Alfresco VRE as a necessary and important step to a much bigger project. The researcher of this study agreed with him that the current Alfresco VRE was very limited and did not make provision for processing or computing and storage of huge data sets. This does not, however, mean that all VREs are as limited. The 'bigger project' that the A-VRE-M described could easily be included in a VRE system, which included all the needed functionalities he had mentioned.

(ix) Migrate the VRE to another software platform

The VRE-D mentioned that there had been a request for the VRE system to have data processing abilities, which could perhaps in the future lead to the migration of the data on the current Alfresco platform to another platform that will have that functionality.

(22) Has the use of the VRE benefitted your research / work processes? If so, how? If not, please explain.

In 6.3, it was mentioned that this question was asked to establish the value(s) that the VRE had for the respondents.

A-R1 mentioned that the VRE had enabled her to access her data when she needed it, and the fact that it synchronised her data, ensured that her data were kept secure. This corresponded with Carusi and Reimer (2010: 18-19), and Neuroth, Lohmeyer and Smith, (2011: 225), mentioned in 5.3, who stated that a VRE can provide a platform where researchers can secure the short-term storage of their data, and a safe place where they can save their data directly. The access-to-data aspect was further mentioned by Yang and Allan (2010: 68), and was cited in 5.3.

A-R2 shared that the system had not directly benefitted her research process, but it had made it easier to share documents within the group. The sharing of data aspect was also mentioned in 5.3 and a number of authors were cited on this aspect of data sharing, for example Carusi and Reimer (2010: 19, 20), and JISC VRE (2006). Filetti and Gnauck (2011: 237) were of the opinion that data sharing is the key element in a VRE. A-R2 was also of the opinion that it should make the A-VRE-M's life easier, because he had access to the files of all the members of the group (Case Study A).

A-R3 indicated that the system provided a safe space to store her data, and allowed for the sharing of data, and that this had been quite valuable. The shared folder with the Liquid Nitrogen records had also been quite useful. A-R4, however, did not find the use of the VRE beneficial. The only value that the VRE had to him was that the backup of data gave him peace of mind. A-R5 indicated that she benefited somewhat by using the VRE. As examples, she mentioned the shared file function, and also revealed that she found the uploading to the VRE easier than onto a cloud. The A-VRE-C indicated two benefits of the VRE: the ability to backup data, as well as in situations where data had been lost and could be retrieved again via the VRE. She foresaw that the VRE would eventually develop into a type of virtual lab book. The A-VRE-M indicated that he knew that their data were safe and backed-up, and that they could access the data whenever they needed it. He again mentioned the need for a system that could help with the computing power and storage capacity they would need for the huge number of data sets they will be working with. At the time of the interview, the VRE had not been meeting that need.

The concept of backup of data as mentioned by A-R4, A-VRE-C, and A-VRE-M were not mentioned directly by authors in the literature (see 5.3), and was therefore a valuable contribution to the study.

The answers from respondents in Case Study B (B-R1, B-R2, B-R4, B-VRE-M), in contrast to Case Study A, were more positive, probably because the current VRE platform, Alfresco, met more of their needs. They also used more of the functionalities in the VRE. B-R1 mentioned that Alfresco was easy to use and was even better than Dropbox. In Dropbox, one could not edit things, but with the VRE one did have that functionality. B-R4 stated that the VRE had made life much easier. She then mentioned how difficult it had been to find something in e-mails, but with Alfresco, one dealt with your own research data and nothing else. This contrasted with what respondents (A-R1, A-R2, A-R3, A-R4, and A-R5) answered in question 23, namely that the Alfresco VRE system was not as user-friendly as it could be. Everything, according to B-R4, was in a central space, and was in the right order. The feedback and workflow function had been a very valuable feature to her. She then expressed the need to be able to store data long-term there. The B-VRE-M mentioned that the VRE had assisted them in organising their research more. It had also given her a better overview of where the student researchers were in terms of their research, and what they were doing. The only respondent in Case Study B to express some uncertainty about the system was B-R3, as she had not used the system yet and still needed to get to know the system.

Concepts mentioned by respondents in Case Study B, which were not found directly in the literature study in Chapter 2, were the functionality of a VRE being a central space, the provision of a workflow function, the feedback mechanism, and long-term storage. These were valuable contributions to the study.

(23) Were there any obstacles in using the VRE? Would you suggest any changes?

The members of both case studies identified the following obstacles:

(a) The fact that other members in the group did not use the VRE to its full potential

A-R1 described herself as more technologically astute and therefore did not experience the VRE as complex, neither did she find any obstacles in the system. She did, however, describe the fact that other people in the group did not use the VRE system to its full potential, as an obstacle. She gave examples of the workflow function and the sharing of things, which did not function as they should because others in the group were not using these. She also mentioned that some members of the group had not used the VRE as often as they should, which led to people forgetting how some of the features worked. These people then fell back on old methods that they were used to. The suggested a change in the synchronising function. She indicated that she would like to see that the system automatically synchronised when she switched on her computer, and if she had done changes to a document, that the system automatically versioned it. This would be valuable, as a number of the members had not been backing up their data as regularly as they should. The system, according to her, could also be made more user-friendly for people that are technologically challenged. A-R2 argued in the same line as A-R1, and mentioned that the system itself is fine if one had attended the training sessions. She also found the fact that members had not been using all the functionalities in the VRE, as they should, as an obstacle. She mentioned as examples, things that are more geared towards interaction, sharing, and workflows. She indicated that she thought it had to do with the group's culture.

(b) Computer illiteracy

A-R3 indicated that she was not very computer literate, and that had been an obstacle at first, but the more she used the system, the easier it became. The B-VRE-M also indicated that it was only her own computer illiteracy that was an obstacle in her use of the VRE.

(c) The VRE system was not user-friendly and intuitive

A-R4 felt that the VRE system had not been very user-friendly. He suggested adding a function that would automatically synchronise his data when he switched on his computer. He did not want to first login to the system. He also wanted the system to be much more integrated with his work processes. He was of the opinion that he should be able to do most of his research processes, such as writing papers, protocols etc., on the system. A-R5 found the VRE system very user-unfriendly and difficult to navigate. This could perhaps be because she had not attended the full training sessions that the rest of the respondents had undergone. The A-VRE-C mentioned that the VRE system had not been as 'smooth' as she would have liked it to be. As examples, she mentioned that every now and then they encountered a small hurdle that they had to attend to, and sometimes some of the functionalities were set in a specific way, but they would want it in a different way. This had made it difficult for her to manage, but she did try to solve these problems as they went along.

The A-VRE-M stated that they needed computing power and huge storage space. They had also been experiencing a number of bottlenecks, for example, with the management of the data. He expressed his concern that although the management of the data should be done properly, it should not be done in a top-heavy manner. In other words, it should not hamstring researchers, but should rather make it easier for them. This means that the VRE system should not be unnecessarily complicated, but rather intuitive.

B-R1 stated that she at first had to understand what 'sites' meant. She could only access her information when she went into the 'site'. She had also found that the system was not as intuitive as one would want it to be. This became evident when it proved to be difficult to re-orientate herself on the system and its functionalities, after not using the site for a period of time. B-R2 indicated that the system had been easy for her to use, but stressed that it could prove to be difficult for someone that has not been using a computer that often. B-R3 had, at the time of the interview, not used the system much yet, but mentioned that one needed to remember one's password, and needed access to the Internet. She had been very impressed that the system could handle uploading of big files in a short time. She also liked the social aspects of the system.

(d) Problems with Internet browser and Internet speed

B-R4 experienced problems with her Internet browser that did not want to open Alfresco. This was solved when she upgraded her browser. She had also experienced slow Internet speeds, which made it difficult to upload files. This was solved when she upgraded her Internet at home. She mentioned that she would like to see the adoption of Alfresco across the University.

(e) Login-problem

The B-L mentioned that she had only experienced a login problem once, which was quickly sorted out by the VRE-D. She was very happy with the system and was impressed by the fact that she could have it on all her devices and could access it from anywhere.

7.3.1.2 Questions To VRE Managers

The VRE Manager was listed as a human component under the core group in 3.5.7.1.

(24) How and when did the VRE project to which you belong, start and develop?

The A-VRE-M mentioned that the VRE-D as well as the researcher of this study, together with a senior staff member from the Department of Library Services, met with him and his research group to discuss the possibility of starting a pilot project on RDM with his group (see 7.2.1.1 and 7.2.1.2). Following this, more meetings were held with him and the group to discuss their needs and requirements (see 7.2.1.3 and 7.2.1.4). During these meetings, the research group (Case Study A) were very specific about things such as confidentiality and securing their data. The first platform to pilot the VRE was based on Moodle software (an open source LMS) (see 7.2.1.3, 7.2.1.5). This was then replaced by a VRE based on Alfresco Software (open source Enterprise Content Management System) (see 7.2.1.8). This research group (Case Study A) also acquired their own dedicated server on which the Alfresco platform could run. The VRE on Alfresco has since then been developed by trial and error and by learning in process.

The B-VRE-M stated (as mentioned in question 6) that the VRE in Case Study B was started after the she had heard of the potential for the usage of the VRE by Case Study A (see 7.2.2.1). Alfresco was deemed by the B-VRE-M as a good system for the group to use as a VRE tool. In other words, Alfresco was already the system of choice, when the group started.

(25) What are your tasks as VRE Manager?

The A-VRE-M described his role as ensuring that members of the VRE stay engaged in the VRE, by sending reminders to members on a regular basis, via e-mail. He also encouraged members to upload their data onto the VRE.

In Case Study A, the VRE Champion was A-VRE-C, listed in 3.5.7.1 as a human component under the core group. The interviews revealed that the A-VRE-C also performed a VRE Manager function. The A-VRE-C indicated that when new students joined the group, she created a user profile for them on the system, and invited them to join the group on the VRE. She also disabled their access when they left. In other words, she controlled the authentication and levels of permissions, which corresponds with the VRE facilitator role under human components, mentioned in 3.5.7.1. She also acted as a point of help, in other words, when people needed help, they would send her an e-mail, and she then tried to sort it out. If she found that she could not help, or that the problem proved to be too complicated, she contacted the VRE-D via e-mail, which correlates with the liaison role of the VRE facilitator mentioned in 3.5.7.1. The B-VRE-M indicated that her role was that of supervisor to the student researchers. She stated that she participated, where possible, in a small manner in their data collection. She also monitored their data analysis and write-ups.

(26) How do you ensure that the members stay engaged in the VRE?

The A-VRE-M admitted that he probably had not done this very well, but then mentioned that they had discussed this fairly frequently. He also reiterated that he sent the student researchers reminders on a regular basis. He stressed, however, that it might be a good idea to institute a formal process to remind everybody. He then mentioned that, in cases where some of his research students had to deal with big data, he encouraged them to upload it onto the VRE. The A-VRE-C indicated that she had found this very difficult. When new researchers joined the group, she demonstrated to them how the system worked and how they could upload their data, etc. She stressed further that people did not tend to take care of their data, and only after they lost data, did they realise its importance. She had also been sending regular reminders to the members of the VRE to upload their data. The B-VRE-M mentioned that she had been encouraging the members in the group on a continuous basis to use the VRE to upload their data. She then expressed that she might need to create a structure on the VRE and then compel them to upload to it, in line with certain requirements.

The role of the VRE facilitator mentioned in 3.5.7.1 could, in other words, be expanded to include encouraging or coaxing members to use the VRE on a regular basis, training of new VRE members on the functionality of the VRE, as well as formal agreements to compel the members to use the VRE. This, however, should not be necessary if the system is user-friendly enough. The VRE system should actually be of such value, and so easy to use, that members of the group would feel that they cannot do without it.

(27) How do you handle technical problems that surface in the VRE? Give examples of how such problems were addressed.

As mentioned in 6.3, this question was asked to determine if technical problems in the VRE are fed back to the VRE designer, and if the VRE was adjusted accordingly. The A-VRE-M indicated that the technical problems that arose in the VRE had been handled by the A-VRE-C. The A-VRE-C confirmed this, and mentioned that if a problem surfaced, she generally tried to solve it, but if she could not solve it, she escalated it to the VRE-D. She gave an example of a problem one of members of the group was having with synchronising (see also 7.2.1.10). One of the respondents had tried to synchronise her files to the system, but the system generated duplicate copies of the files. This specific respondent did not select the files she wanted to synchronise, with the result that the system synchronised her whole VRE instance to her computer, resulting in duplicate copies. This was something that could have been solved if the settings in her VRE instance had been set up correctly. Another example she mentioned had to do with deleting files on the VRE. One of the members of the VRE had deleted something on the VRE, but then through synchronising, the system deleted this also on this person's computer. The instance of the system had been wrongly set up to do a two-way synchronising. Fortunately, she could salvage the file from the trash function in the system. These types of problems caused the members of the group to be very hesitant to use the VRE.

The B-VRE-M indicated that, at some stage (see also 7.2.2.7), she had encountered a problem where the system was not sending any notification e-mails to her – a function of the VRE system that normally notified members of the VRE when something on the system had been updated, or when someone had uploaded something on the system. She immediately contacted the VRE-D and he corrected it speedily and solved the issue.

The way the A-VRE-C and B-VRE-M had handled technical problems by contacting the VRE-D, was in line with what was found in literature as discussed in 3.5.7.1, where it was mentioned that part of the role of VRE facilitator (VRE Manager and/or VRE Champion) was to liaise with the VRE designer.

(28) How do you address additional needs in the VRE? Give examples

The A-VRE-M indicated that the A-VRE-C handled additional needs that might arise in the VRE. The A-VRE-C confirmed this and mentioned that if someone in the group asked for an addition to the VRE, she would e-mail the VRE-D, and then he would indicate if this was feasible or not at that time, or if this might be possible at a later stage. She stated that they had not, however, received many of these types of requests, and she could not think of an example at the time of the interview. The B-VRE-M mentioned that she would either contact the VRE-D or the researcher of this study to find out if something could be added to the VRE. She also could not mention an example. By contacting the VRE-D for additions to the VRE system, the A-VRE-C and B-VRE-M again fulfilled their role of liaising with the VRE designer as mentioned in 3.5.7.1.

(29) How do you define the added value of the VRE for the members of the VRE?

In Question 22, the A-VRE-M indicated that he knew that their data were safe, and backed-up, and that they could access it when they needed it. The A-VRE-C mentioned that in her opinion, the VRE added value to the students' research processes, because it enabled them to back-up their data in a safe place, where their data could not be lost. This aspect of using a VRE for backing up of data was not found in the literature specifically, but the related concept of storage of data is listed, however, as an action that takes place in the proposal stage and the experimenting and analysis stage of the research lifecycle in 5.2 and in Table 5.1. The A-VRE-C expressed further that it would have been great if the VRE system could replace their paper-based processes totally, but for that to happen, it would have to be integrated with the computers in the laboratory. The B-VRE-M indicated that the moment they obtained a license, they could upload the raw data into a repository, which typically would happen in the dissemination of findings stage as mentioned in 5.2. A data repository was also listed as an RDM component in Figure 5.2b. The B-VRE-M further mentioned the value of having everything together in one place, and that by having the data there, would also help in gaining international standing amongst their publishing colleagues. It is important to note that although Alfresco is a perfect solution for the backing-up of data in a secure environment, it is not a data repository where data are published, and it is closed to persons outside the research group. It is also not a solution that can be used for the long-term storage and curation of data.

(30) How do you ensure quality control in the VRE?

The A-VRE-M mentioned that he did an inspection of his student researchers every now and then, to see what they had been uploading, but found it too time consuming to do quality control of everything. He therefore did not spend too much time on that. The A-VRE-C stated that she had not been aware that she was supposed to be doing that. She was of the opinion that the responsibility for quality control lies with the student researchers themselves. She also mentioned that it would be too time-consuming for her to check the quality of everything that was being uploaded onto the system. She suggested that a way to ensure quality control in the VRE would be to create specific folders where data could be uploaded. The student researchers would then take the responsibility for deciding which data could be uploaded, because they knew their data better than anyone else. She also mentioned that it would be very difficult for her to do quality control of every student researcher's files/data, because of the diverse research areas they had been working in.

The answers received from the A-VRE-M and A-VRE-C indicated that this is an area where Case Study A had been lacking, and although the responsibility for checking the quality of the data that were uploaded onto VRE, seemed to have been placed on the student researchers themselves, there seemed to be no guidance in this area. This could potentially become a critical problem when the volume of data uploaded onto the VRE increase exponentially.

In Case Study B, the B-VRE-M indicated that she had been engaged in the data collection procedure, and had monitored the whole process to ensure quality.

(31) Does the project have a formal RDM strategy / plan? If yes, please elaborate. If not, please explain.

As mentioned in 6.3, this question was asked to establish if a RDM strategy/plan was in place to structure and guide RDM activities in the VRE, and also to determine what it entailed. Answers received from respondents in both case studies revealed that no formal strategies or plans existed.

The A-VRE-M revealed that Case Study A did not have a formal RDM strategy or plan in place for their group. He had, however, discussed their vision and strategy in Question 19, where he indicated that he saw the VRE as just a point along the line of where they were heading. He also indicated that they need fluidity to operate and that a formal strategy or plan might be counterproductive at this early stage in the VRE. The A-VRE-C confirmed what the A-VRE-M had said, but mentioned that the members of the group had an unwritten understanding that the A-VRE-M wanted them to store their data on the VRE and that the data should be available for retrieval. In a similar vein, the B-VRE-M indicated that the group (Case Study B) had an understanding amongst themselves of what was expected. The group, however, also did not have a formal strategy or plan.

(32) If the project does have a formal RDM strategy / plan, how do you ensure compliance within the group?

As indicated in Question 31, there had been no formal strategy or plan in Case Study A, with the result that no compliance could be checked at the time of the interviews. In Case Study B, the B-VRE-M indicated that although they did not have a formal strategy or plan in place, she checked that the student researchers uploaded onto Alfresco.

(32)(b) Your group do have a librarian that is helping. What role do you see the librarian play in the VRE?

The researcher of this study added this question, which was only directed at the B-VRE-M, because Case Study B was the only case study that had a librarian as VRE member. The B-VRE-M indicated that the librarian co-searched the journal databases for relevant information for the individual members (the student researchers) of the VRE. Her view of the role of the librarian was very limited compared to the description of her role by the librarian (B-L) herself in the answer to Question 46.

7.3.1.3 Questions To The VRE Designer

(33) What are your tasks as VRE designer?

The VRE Designer / Developer was listed as a human component under the peripheral group in 3.5.7.1, where it was mentioned that a VRE designer would need access to all levels of the VRE to develop, build and sustain its features. The VRE Designer / Develop component was also shown in Figure 5.2b.

The purpose of this question was to get more clarity on the role of the designer in the VRE and to list his/her responsibilities.

The VRE-D described his daily tasks as follows:

- See to it that the servers were up and running (this was monitored automatically through e-mail notifications that he received);

- See to it that backups of the system had been completed (this was monitored automatically through e-mail notifications that he received);
- Keep up to date about the latest versions of Alfresco, and about possible plugins to the system (he tried to keep all Alfresco installs on the same version);
- See to it that the VRE system was performing well, was quick and accessible, visually appealing, and made sense, and that it was easy to navigate through the system;
- Make changes to the VRE system if there were requests for specific features, and if this feature(s) was a default feature, he just enabled that, but if a customisation needed to be done, he would do that in a test environment, and when ready, add that onto the live VRE environment;
- Register users on the VRE and set levels of permissions on who can access what;
- Train the users of the VRE system;
- Act as a consultant to members of the VRE regarding any problems they encountered; and
- Scan the environment for other VRE systems.

(34) What is the process you followed to design these two VRE projects? Did you first create a prototype for the VRE(s)?

As mentioned in 6.3, the purpose with this question was to establish what steps were followed in the design process in order to construct the final product.

The VRE-D indicated that his first introduction to VRE software was an earlier version of HUBzero, which had not been very user-friendly. After that he became involved in the CSIR's Natural Products VRE, and this provided him with a framework to base all his future development and testing on. He also investigated Moodle, Sakai, Chisimba, and at a later stage, Alfresco, as possible software tools to use in the development of a VRE. He tested each one of them to see which would be the easiest to customise. Initially he found Moodle to be the easiest, and this led to the first VRE prototype for Case Study A, which was built using Moodle (see 7.2.1.4 and Figure 7.4). He then discussed the different technologies on which these VREs were designed and stated that most of the VREs he had dealt with, were developed on open source platforms that had good community support, and were based on Java, PHP, MySQL or PostgreSQL. Community support ensured that help was available when guidance was needed with regards to resources or specific coding. As mentioned, Moodle was found at first to be the easiest platform to work with, because it was based on Java, PHP and MySQL, and he had knowledge of these. He also indicated that, throughout the design of the VRE, he had been in engagement with a senior member of the Department of Library Services at the University of Pretoria, as well as with a senior member from the CSIR, to get their input on what a VRE and RDM entails. The Moodle instance that was originally introduced was subsequently replaced by an instance of Alfresco (second instance) (see 7.2.1.8).

At the time of the interview, the VRE-D was investigating HUBzero as a potential replacement for Alfresco (which could perhaps be a third instance). He indicated that Moodle, which is essentially a learning-based platform, was initially chosen as a suitable platform to teach people how to work with their data. Some of the shortcomings that he encountered with Moodle were that document features such as technical metadata did not exist. Moodle also did not have a central repository where files could be indexed, searched, tagged, categorised, or where metadata could be added – all very important aspects to managing research data in a VRE. While Moodle was running, he decided to investigate systems to address these shortcomings. This led to the introduction of Alfresco. Alfresco was an open source system, and he found it to be quite strong in terms of enterprise content management, and even stronger than some proprietary systems. Following this, he investigated the possibility to simulate on Alfresco, what Moodle could do. The results showed that Alfresco could do all the core features found in Moodle. In addition, it had a very strong document management system built into it. This document management function is something that is needed in terms of managing one’s data, and proved to be decisive. The Alfresco system allowed for versioning of files (documents or data), and it also made provision for audit trails of data and metadata, etc. HUBzero, which the VRE-D had been investigating at the time of the interview, was found to be specifically designed as a VRE tool, by Purdue University in the USA. The VRE-D mentioned that he had tested HUBzero during a workshop with a group of students that had attended a Carnegie-funded Continuing Professional Development (CPD) course presented at the University of Pretoria. He described HUBzero as the best of Moodle and the best of Alfresco built into one system. In addition to these features, it also had sequencing software built into it. In other words, HUBzero also included the processing part of the research process. The VRE-D envisaged a natural transition from Alfresco to HUBzero, and expected that it would be quite easy to transfer the case studies onto HUBzero. He felt that it would be counter-productive to try and customise Alfresco to be able to do what HUBzero could do (in other words, try and reinvent the wheel), if HUBzero already had the necessary features built into it.

(35) What software(s) did you use to design the VRE?

The VRE-D stated that he had experimented with VREs that were developed using open source software, e.g. Moodle, Chisimba, Alfresco, and HUBzero. These platforms, according to him, were based on Java, PHP, MySQL or PostgreSQL.

(36) How did you decide upon the specific software(s)? Why did you use these specific software(s)?

The reasons why the VRE-D chose these different types of software components were discussed in Question 34.

(37) How did you determine which functionalities (components) the members of the VRE groups needed in the VRE?

The VRE-D stated that they (he, the researcher of this study forming the design team, as well as a senior member of the Library Services) had a discussion with the group (Case Study A) to determine their requirements (see 7.2.1.4 and 7.2.1.5). The VRE system was then set up to meet these requirements. In Alfresco, the list of features and functionalities that are available can be quite overwhelming, and he indicated that after this discussion, he had removed some of these features and functionalities. The features and functionalities that were left were in line with the requirements of the group. He mentioned that Case Study A also had a custodian (A-VRE-C) for the system, and that she had indicated some of the features that they would want, and those they would not want. He further reiterated that members in Case Study A requested some features after they had started using the VRE platform. They requested, for example, a Calendar function and a discussion function. They also, at some stage, requested automated backup of their instrumentation. He had a meeting with them to discuss this, and they, at the time of the interview, were just waiting for a network project to finish, after which they planned to link all the electronic instruments and equipment in the research laboratory (see 7.2.1.14) to the VRE platform. This would then make it possible to transfer backups onto the VRE from these instruments and equipment. The VRE-D mentioned that, similarly to Case Study A, they had a meeting with the members of Case Study B to discuss their needs. This group requested to have the ability to view videos within the VRE. The VRE-D then installed a module on the VRE so that they could preview videos within the VRE. They also indicated that they would want a survey tool in the VRE, and the VRE-D then installed an X-frame into the VRE that would enable surveys (see 7.2.2.4 on the plugin of a survey tool).

(38) What are those functionalities that you made provision for?

The VRE-D indicated that data storage was the first functionality that had to be provided for. Each individual had his/her own working directory that only he/she, the A-VRE-C and the A-VRE-M could see. The group also had a shared library where they could share the data that everyone was supposed to see, for example their standard operating procedures (SOPs) for the laboratories. There was also a folder for completed research projects where they had uploaded retrospectively, projects of students that had finished their studies and had left the University. The system furthermore had a tools and software function where they could gain access to the synchronising function and other standard Alfresco tools. Another feature was members' lists, indicating who the members of the various sites were. Respondents had been encouraged to complete their profiles, which would enable others in the group to see who was busy working on the site. The site calendar had been activated, but as mentioned earlier in Question 8, had not been actively used by members.

The system also had a workflow function and the VRE-D indicated that he had given training to the members of both groups, on how workflows operate. A Workflow function was something the respondents from Case Study B identified in 7.2.2.4 as a real need. Another feature that had been enabled by the VRE-D was the Google Maps Integration, which showed the geographical location where photos were taken. The VRE-D also mentioned that the document management side of the system had a whole range of features, for example the offline capabilities, versioning, editing in Google Docs, etc.

(39) What are the hardware and software infrastructure specifications for RDM activities within your VRE (e.g. storage and computing capacity needed)?

The VRE-D revealed that Alfresco was, at the time of the interview, running on three servers, which all replicated each other (see 7.2.1.14). The first server had been purchased by Case Study A, and was also situated at their site on one of the University's satellite campuses. The server was a Dual CPU Xeon server with 16 GB RAM, and had a 3 TB RAID (Redundant Array of Independent Disks) configuration for redundancy (or in case of disk failure). Should a drive fail, one could just replace the drive. The second server was situated in the Bio-Informatics Department on the Hatfield Campus of the University. The VRE-D had not been certain about this server's total capacity. The third server was situated in the Merensky II Library Building, also on the Hatfield Campus of the University (see 7.2.1.14). This server was a 16 Core server, with 32 GB of RAM. The Alfresco instance for Case Study B ran on the third server in the Merensky II Library, with 2 TB of storage dedicated to it, and had been replicated on the other two servers. All three of these servers were linked to the University's backbone / network, which was quite fast in terms of data transfer. All of these servers furthermore were running Linux Open Source, and all of them were running the same versions of Alfresco.

(40) How did you ensure that the data in these VRE's are protected from loss or damage?

In Question 39, the respondent (VRE-D) indicated that the data from these two VRE's had been replicated on three servers geographically spread out on the University campuses, to protect data from getting lost or damaged.

(41) Do the VRE systems make provision for data publishing, as well as long-term preservation of data?

The VRE-D indicated that the VRE systems did not make provision for data publishing or long-term preservation of data. At the time of the interview, these processes had to be done manually.

(42) Did you have to make any adjustments to the VRE? If so, what did you do?

The VRE-D mentioned that he did make adjustments in the coding of the system, by changing some of the scripts and the integrations. He created a folder on the system where he had been backing up all the customised code. In the case of upgrades, the file could then be re-imported so that one could have a list of all the customisations. The VRE-D further indicated that some of the changes were configuration changes, while others were custom scripts (Java scripts), which he then edited. The survey tool that was added to Alfresco was an example of such a customisation.

(43) What type of training, if any, did you give to the students, researchers, librarian, and VRE Managers?

The VRE-D stated that he had held official training sessions with members of both case studies. He first presented two group-training sessions for the members of Case Study A on the Hatfield Campus of the University (see 7.2.1.6 and 7.2.1.9). One of these training sessions had also been attended by the B-VRE-M (see 7.2.2.2). The VRE-D had more than one group-training session with the members of Case Study B, on the campus where that group was situated (see 7.2.2.5 and 7.2.2.10). The content of the training was very basic and consisted of the following: how to access the system; how to log onto the system; one's personal dashboard; the site dashboard; how to join a site; where to find one's directory on the site where one's data are stored; where the common directory was, and where that data were stored; how to drag and drop; how to do versioning; how to download a file; how to delete a file; how to create a folder, etc. These training sessions, according to the VRE-D, took about an hour and a half per session. The VRE-D further indicated that he had also conducted training sessions with certain individuals that were not able to attend some of the group-training sessions. He furthermore had follow-up training sessions with individuals that did not understand things well during the first round of group-training sessions.

(44) What future developments do you envisage for the VREs?

The VRE-D mentioned that he would like to see that HUBzero was adopted as the platform for the VREs at the University. He also indicated that he would want people to get involved in developing their own tools for the VRE. He would furthermore like to see HUBzero as part of the University's authentication directory, for example the active directory, so that any researcher can sign onto the VRE. In other words, researchers at the University would not have to login to their instance of the VRE, but would already be logged in through their system authentication and could use the functionalities of the VRE without having to login again on the system. He also foresaw specific units, or faculties developing specific tools. Other functions he foresaw were the capability of publishing directly from the VRE onto a repository, as well as the long-term preservation of the data in a data archive / preservation storage solution.

7.3.1.4 Questions To The Information Specialist / Librarian

(45) Do you think a librarian has a role to play in a VRE? If so, what do you see as the potential role(s) a librarian can play in a VRE?

The B-L indicated that she thought a librarian had a role to play in a VRE, but that it was dependent on whether the members of the VRE allowed the librarian to be part of the VRE. She then described her role in the VRE as collecting information at the onset, collecting data, and assisting the researchers in formalising their protocols. The VRE also provided her with a platform to communicate with the researchers. She furthermore assisted the researchers in organising their files and folders. Her answers were very much in line with Bowers and Van Deventer (2012), mentioned in 3.5.7.1, who see the role of librarians more in terms of populating the VRE with content, as well as structuring access to the content.

The B-L, however, did not mention the following tasks that were identified in the literature study in 3.5.7.1:

- Ascertain the "user requirements and facilitate user evaluation" for the design of the VRE through the trusting relationships and liaisons they have with researchers, and share this with the VRE designer(s) (Wusteman, 2010: 69);
- Create tools and interfaces that will allow for the searching and usage of the information resources (Candela, Castelli and Pagano, 2009: 248);

- Conduct e-Research literacy training by training researchers to use and manage VREs, as well as the tools within them (Wusteman, 2010: 69);
- Ensure that the appropriate information-related standards and solutions are used in VREs, especially with regards to the usage of metadata (Wusteman, 2010: 69);
- Check “that open access publications do not violate any third-party rights before publication,” and advise researchers on copyright, open access and licensing issues (Carusi and Reimer, 2010: 54, 73);
- Collect, curate, preserve, maintain and archive various digital assets such as software repositories, research workflows, and research outputs (publications) (Candela, Castelli and Pagano, 2009: 249).

(46) What are your tasks as librarian in this VRE?

This question links up with Question 45, where the B-L mentioned some of the tasks she performed. In this question, she elaborated more on the tasks that she performed. The collecting of information, which is part of the traditional librarian role, consisted of literature searches that she performed for researchers, which she then uploaded onto the VRE. The researcher(s) would also do literature searches and then compare it with the results that she as librarian had found, in order to ensure that they had covered the whole spectrum of literature on a topic. She also gave guidance in organising the literature in folders and files.

(47) Do you have any specific role in terms of RDM in the VRE? If yes, what? If not, why not?

The B-L felt that she did not play a specific role other than what was mentioned in Questions 45 and 46. She felt that because the VRE was still in a pilot stage, she was not certain of the potential role(s) she could play in the VRE.

(48) How did you get involved in this VRE(s)?

The B-L indicated that she got introduced to, and involved in the VRE, through the B-VRE-M.

(49) What do you see as the value of a VRE for you as librarian?

The VRE, according to the B-L, had been a much more regulated and organised environment, where everything (for example communications and uploads) was in one place. Another benefit of the VRE was that members could access it from anywhere. She could upload something on it, while members did not need to be at home or at the University to access it, but could access these files from anywhere. The B-L also mentioned that the librarian could create metadata for the files that the researcher had uploaded. The researcher of this study then asked the B-L if she had helped the members of the VRE with file naming conventions, but the B-L indicated that she had not done that.

7.3.2 Summary Of The Summative Evaluation

The summative evaluation, as mentioned in 7.1, consisted of semi-structured interviews with the members of each of the two case studies, as data collection method. The answers to these questions were then mapped to findings in literature and the findings of the formative evaluation.

In 7.3.1.1, where the questions were directed to the postgraduate student researchers, the results showed that VREs developed through various stages, and that they consist of a number of components that make them successful. For example, the human components involved, consisted of the following role players:

- Role players before implementation included a member of the University Executive (the Vice Principal Research), members of the Library Executive (Library Director, and a Deputy Library Director), the Library Advisory Committee, a Dean of one of the faculties, the chair of the Ethics Committee of that faculty, the head of one of the institutes of that faculty, the researcher of this study, the repository manager, and the designer of this study.
- Role players during the development and implementation of the VRE comprised of the student researchers involved, a VRE Manager, a VRE Champion (in Case Study A), a librarian (in Case Study B), as well as a design team.

The answers received from the respondents in each of these case studies showed the importance of providing comprehensive training to each of the members of these VRE groups. Those who did not attend the comprehensive training sessions had difficulty in navigating the VRE platforms. Those who had joined the VRE groups at a later stage also experienced difficulty in navigating the VREs. The respondents in each of the case studies knew what their perspective roles were in these VREs, and the VRE-D assigned each a level of authentication according to their roles.

Results from each of the case studies showed that there was a need for the provision of multi-disciplinarity in the VREs, and this in turn had an influence on the different ways each of these VREs and their components developed. The majority of respondents felt that their particular VREs were developed around their topics and were not specifically driven by the technologies themselves, which is a positive outcome, as technology-driven projects tend to be too prescriptive. Fortunately, most of the respondents, at the stage of the interviews, had a broad understanding of what a VRE constituted. The answers received were perfectly aligned with results found in the literature. The respondents' answers showed that some saw it as an online or digital system/framework that is cloud-based.

Most of the respondents indicated that they were afforded the opportunity to give input in the design of the VRE, which is in line with what was found in the formative evaluation in 7.2.1.3, 7.2.1.4, 7.2.1.7, 7.2.1.10, 7.2.2.3, 7.2.2.7, and 7.2.2.11. These inputs were not always from all the members of these groups, for example in Case Study A, the A-VRE-C had a big role to play in providing input and in Case Study B, the B-VRE-M had a big role in providing input, which led to some of the members feeling that their input in the design of these VREs was minimum.

During the interviews, the members of the VRE listed a number of things that a VRE could be used for:

- Storage or archiving;
- Provision of access to information and data, and allowing for the sharing of data;
- To facilitate collaboration and interaction;
- Management of data; and

- Back-up of data files;
- Usability across organisational boundaries; and
- Possibility to add plug-ins to the VRE.

With regards to the components of the Alfresco VRE platform, some of the respondents indicated that a VRE could have multiple components. The answers received from respondents, however, showed that not all of the components were utilised. The components used by the members, those not used by them, those added to the VRE, as well as those components that would need to be added in future, are indicated in Table 7.7.

Table 7.7: VRE Components

Components/Functionalities of the VRE				
Component	**Used**	**Not used**	**Added**	**Needs to be added at a later stage**
Create a site	✓			
Edit Your profile	✓			
Search	This was used by some of the members in both case studies.			
Site calendar		✓		
My discussions	Some of the members of Case Study B used this.	This was not used by members of Case Study A.		
Following	This was used only by the A-VRE-C and the B-VRE-M to monitor members' activities.	The majority of the members of both of the case studies did not use this.		
My Files (Drag & Drop, upload files, create folders)	✓			
My Activities (News)		None of the members in either of the case studies used this because the groups were small enough to follow activities, without having to use this.		

Site Activities	✓			
My Tasks (Workflow Function)	This component was very important for members from Case Study B.			
My Documents (Keeping track of own content)	✓			
Shared Files (Files everyone has access to)	✓			
People Finder		✓		
Invite Users	Only the A-VRE-C and B-VRE-M had rights to use this.			
Discussions		✓		
Document Library	✓			
Categories	This was used by the majority of members from Case Study B.	This was not used by the majority of members from case Study A.		
Tags		This was not used by the majority of members from Case Study A and B.		
Favourite		This was not used by the majority of members from Case Study A and B.		
Like		This was not used by the majority of members from Case Study A and B.		
Comments	Half of the number of members from Case Study B used this.	This was not used by the majority of members from Case Study A.		
Share	The majority of members from Case Study B had used this.	Four of the members of Case Study A had not used this.		
Edit Properties		The majority of members in both case studies indicated that they had not edited the		

		properties of their files.		
Edit Offline	The majority of members in Case Study A used this.	The majority of members in Case Study B used this.		
Dublin Core Metadata Template		All the members from both case studies have not used this.		
Manage Permissions	The A-VRE-C, the B-VRE-M and VRE-D had rights to use this.			
Upload New Version	In each case study, three members indicated that they had been using this.	In Case Study A, four members and in Case Study B, four members indicated that they had not been using this.		
Download Function	✓			
Instrument Backups		✓		✓ A tool that can capture / generate data from instruments.
Software Backups		None of the members of Case Study A used this, and the majority of members of Case Study B indicated that they did not use this.		
Survey or Questionnaire Tool			✓ This tool was added to / plugged into the VRE by the VRE-D, for use by members from Case Study B.	
Publishing Function		✓ This component only allows publishing to social media sites.		
Mobile Syncing with Alfresco	Only one member of Case Study A used this and only	The majority of members in both case studies		

	two members of Case Study B used this.	mentioned that they had not used this component / function.		
Desktop Syncing with Alfresco	The majority of members in both case studies used this.	Two members in Case Study A and two members in Case Study B did not use this.		
Analysis software (e.g. analysis software for systematic literature reviews, and an electronic movement analysis system such as Qualisys)				✓
A tool (component) that can be used to do a simulation of an experiment and in the process, generates data.				✓
Access to data processing programmes within the VRE				✓
The ability to run non-wet laboratory experiments within the VRE				✓
A tool (component) that would be able to generate visualisations				✓
The ability to publish on a data repository by using the VRE				✓
A link to a referencing system, for example				✓

EndNote or RefWorks				
The management of data was viewed by both case studies as an important component of a VRE.	✓			

The sub-components in the 'Document Library' that were more socially oriented, were not used extensively by members of Case Study A, because the social interaction aspect of this group had not developed to its fullest level yet. These were things such as 'Categories', 'Comments, 'Like', and 'Tags', etc.

In 3.5.7.1, the role of the VRE Champion was identified as pivotal to keeping everything and everyone in a VRE together. The answers from respondents, as well as the notes and e-mails, revealed that Case Study A had a member of staff that was the designated VRE Champion for that group, while in Case Study B, the B-VRE-M performed that function.

The descriptions of what the respondents defined as 'research data' showed that in the majority of cases, there was uncertainty or a wrong perception of what could be seen as 'research data'. The respondents in most instances did not make a distinction between various types of data as listed in 4.2.1, for example research data, referencing data, funding data, collaboration data, and administrative data, but rather grouped most of these types of data under research data. Many of them also included research outputs flowing from the research, as research data. These respondents were, not surprising then, of the opinion that all these different types of data should also be uploaded onto a VRE platform.

The respondents' understanding of the concept of RDM was shown to be limited. Most of the members of Case Study A and B emphasized the storage and accessibility of data, but failed to mention the preservation of data or publishing to an open access repository. Only the B-VRE-M and the VRE-D described RDM in terms of the whole RDM lifecycle.

The answers to the question ‘to what extent could data be managed by using a VRE?’ revealed that the majority were of the opinion that the VRE platform was an excellent tool for the management of research data. Respondents were of the opinion that it could be used for the backing-up / saving of data, metadata could be added to the data (although they did not make an effort to do this), and it also provided easy access to the data. One could also tag one’s files, share one’s files, and protect one’s files, and upload different versions of files - aspects that were not found in the literature. VREs furthermore provided researchers with a place where they could secure their data safely. Another advantage that was mentioned was the ability for co-researchers that were geographically spread out, to access data. It also provided the promotor / supervisor with the ability to monitor the data of student researchers. Respondents, however, did not mention the aspect of collaboration around data, which was found in the literature.

The long-term preservation of data was found to be something that lacked in the VRE platform. Answers received from the respondents showed a total lack of understanding with regards to what long-term preservation of data entails. The VRE-D understood what the concept entailed and mentioned that it was possible to customise the Alfresco system for automatic data preservation using, for example, the BagIt specification. At the time of the interviews, he was testing it manually.

After gauging the openness of the data that were managed in both case studies, it was found that access to the data of all the respondents in both case studies had been restricted for confidentiality and ethical reasons. All of the respondents in Case Study A, except one, indicated that there were legal restrictions on their data, while in Case Study B, only the data of one respondent were restricted by legal requirements. In addition, it was found that the data of all respondents in Case Study A were restricted by proprietary or commercial interests, to some or other degree, while none of the data of the respondents in Case Study B were restricted for these reasons. It was also found that none of the respondents had published their data yet, which meant that they had not added a Creative Commons License to their data, which could have placed a restriction on the usage of their data.

The results from both case studies showed that each of these groups had been using different analysis tools that were more characteristic of the nature of the disciplinary areas they focused on. None of these tools were integrated with the VRE platform, however, which means there is scope to add these to the VRE platform in future.

The responses received from members of both case studies revealed that the majority of them used an array of data visualisation tools, and a desire was expressed to include this as an RDM component in a VRE conceptual model. Neither of the case studies had a formal RDM strategic plan, or RDM policy, in place, and none of the interviewees indicated that they had compiled DMPs, but the members of both case studies knew what was expected of them. The RDM facilitators of both case studies felt that at the time of the interviews, it was too early to formalize things. They needed the flexibility to enable the VREs to develop more, before formalising the VREs and including things such as DMPs.

The respondents from both case studies revealed a number of immediate objectives with regard to the VRE. These have been listed in Table 7.8.

Table 7.8: VRE Objectives

Immediate objectives	Future Objectives
• The storage or backup of data. • The recovery and retrieval of, and access to their data (as well as accessing the VRE remotely from another geographical area). • The idea of having a common, centralised space, system or location. • Data sharing. • The ability for supervisors to track their student researchers' progress. • The idea of making protocols to different experiments available through the VRE, to other members. • The preservation of data. • The ability to work together collaboratively on a document. • The ability to secure data. • The ability to keep track of all communications in one place.	The inclusion of: • Tools for data processing. • Tools for analysis of data. • Tools for sequencing of data. • Tools for simulations of data in the VRE. • The ability of the platform to handle big data sets. • More computing power / HPC. • Multidisciplinary research. • The ability to use secondary data created by other researchers. • Infrastructure that is readily accessible and integrated in terms of their daily routines. • Tools for data visualisation. • A more intuitive interface.

The responses from the majority of members from both case studies revealed that most were of the view that there was no time limit to these two VREs. The (in)ability of Alfresco to handle big data sets was mentioned as a possible debilitating characteristic, which might have an effect on the timespan of using the VREs in their current forms at the time of the interviews.

During the interviews, the respondents disclosed the following benefits that the VRE platform had for their research and work processes:

- Access to their data when they needed it;
- Keeping their data secure;
- Possibility of synchronising their data from their desktops to the VRE platform;
- Providing a place for short-term storage / back-up of their data;
- Making it easier to share documents within the group;
- The editing function within Alfresco;
- The fact that everything was in a central space;
- The workflow function;
- Ability of respondents to better organize their research, including; and
- Feedback mechanism.

Obstacles in using the VRE as uncovered by the respondents were:

- Some of the members of the group were not using the VRE system to its full potential, which, for example, caused the workflow function, the sharing of things, and interactions not to function as they should;
- Some members of the group had not been using the VRE as often as they should, leading to people forgetting how some of the features worked;
- The system did not automatically synchronise data when switching on desktop computers;
- The system was not user-friendly for people that were technologically-challenged, in other words, the system was not intuitive enough;
- The system was not integrated enough with respondents' work processes;
- The system did not have enough computing power and enough storage space for huge data sets; and
- Slow Internet speed.

In 7.3.1.2 of the summative evaluation, where the questions were directed at the VRE Managers, the A-VRE-M indicated that he and his research group was approached by the VRE design team (consisting of the VRE-D and the researcher of this study), as well as a senior member from the Department of Library Services, to discuss the potential for implementing a VRE pilot project with the group (see 7.2.1.1). This was followed by more meetings to discuss their needs and requirements. Moodle was first used to pilot the VRE, but was subsequently replaced by Alfresco (see 7.1.2.8). Case Study A also acquired their own dedicated server on which the Alfresco platform could run (see 7.2.1.4). The VRE for Case Study B developed after the B-VRE-M heard of the potential of the VRE for Case Study A. She subsequently attended the second training session that was held for the members of Case Study A (see 7.2.1.9). The B-VRE-M considered Alfresco as a good system to use as a VRE tool, and this tool was subsequently implemented for the student researchers of Case Study B.

Feedback on the roles of the VRE Managers in both case studies showed the following:

- They ensured that members of the VRE stayed engaged in the VRE, by sending reminders to members on a regular basis via e-mail;
- They encouraged members to upload their data onto the VRE;
- They created user profiles for post-graduate student researchers on the VRE system, and invited them to join the VRE;
- They disabled access when members left; in other words, they controlled the authentication and levels of permissions;
- They participated, where possible, in respondents’ data collection; and
- They monitored respondents’ data analysis, as well as their write-ups.

To ensure that members stayed involved in the VRE, the A-VRE-M sent regular reminders to members via e-mail, but he admitted that this was one area that he did not do well enough in. The A-VRE-C, who also performed a VRE Manager function, indicated that she had found this to be very difficult, but mentioned that she had also sent members of the VRE regular reminders to upload their data, and had to encourage the members in the group on a continuous basis to use the VRE to upload their data onto the VRE.

Technical problems in the VRE of Case Study A was dealt with by the A-VRE-C and when a problem arose, she first tried to solve it herself. If she was unable to solve it, she escalated the problem to the VRE-D. Technical problems in the VRE of Case Study B were handled by the B-VRE-M, who indicated that when she encountered a problem, she immediately contacted the VRE-D.

When additional needs arose in the VREs, the A-VRE-C mentioned that she would contact the VRE-D. In Case Study B, the B-VRE-M indicated that she would contact either the VRE-D or the researcher of this study. In both cases, the VRE-D would consider these requests and would then indicate if it was possible or not to address the needs, at that stage.

The answers received from both the A-VRE-M and the A-VRE-C with regards to ensuring the quality of the data that have been uploaded by the postgraduate student researchers, revealed that this was not done because it was seen as too time-consuming, and the researchers' topics were too diverse. The responsibility for ensuring that good quality data were uploaded was placed on the student researchers themselves; however, there seemed to be no guidance in this area. In Case Study B, the B-VRE-M indicated that she had monitored the whole process to ensure quality.

In 7.3.1.3, the researcher of this study directed a number of questions to the VRE designer (VRE-D). The VRE-D listed a number of daily tasks that he took responsibility for, and then described the steps to be followed in the design process to develop the VREs for these two case studies. He indicated that he first investigated Moodle, Sakai, Chisimba, and at a later stage Alfresco, as possible software tools to use in the development of a VRE, and that he tested them for ease of customization. Initially, he implemented Moodle as the first prototype. The Moodle instance was subsequently replaced by a second instance on Alfresco, which had all the functionalities of Moodle, but in addition, had a strong document management system built into it. At the time of the interview, the VRE-D was investigating HUBzero (a tool specifically designed as a VRE tool) as a potential replacement for Alfresco. The VRE-D further mentioned a number of functionalities he made provision for in these VREs, such as:

- Data storage;
- A working directory for each member;

- A shared library with shared folders;
- A tools and software function where members could get access to the syncing function and other standard Alfresco tools;
- Members' lists;
- A site calendar;
- A workflow function;
- A Google Maps integration; and
- Document management features.

The VRE-D, in addition, revealed that the Alfresco VREs were running on three servers, geographically separated from each other, that were replicating to each other. This ensured that the data were protected from damage or loss. These three servers were linked to the University's backbone / network, which was quite fast in terms of data transfer. All of these servers furthermore were running Linux Open Source, and all of them were running the same versions of Alfresco. The interview with the VRE-D further revealed that he had done some adjustments in the coding of the system by doing some configuration changes and by changing some of the scripts and integrations. He also did some customisations; for example, he added a survey tool to Alfresco. The VRE-D further indicated that he envisaged the implementation of HUBzero as the platform for VREs at the University.

In 7.3.1.4, the researcher of this study directed a number of questions at the librarian (B-L) that was involved in Case Study B. When asked about the role of a librarian in a VRE, she mentioned that it would depend on whether the members of a VRE allowed a librarian to take part. In this case, the B-VRE-M introduced her to the VRE. She then went on to describe the tasks she performed in the VRE of Case Study B. These tasks were:

- Collecting information and data, by doing literature searches for researchers;
- Assisting researchers and providing guidance in organising their files and folders, for example using file naming conventions;
- Assisting researchers in formalising their protocols.

The VRE also provided a vehicle for the B-L to communicate with the researchers. Her answers revealed a limited knowledge of the various roles / tasks a librarian could perform in a VRE. The librarian described the value of the VRE as the fact that it was a much more regulated and organised environment where everything was in one place. The fact that she could upload something on it, while members could access it and download it from anywhere, was also of great benefit.

7.4 SUMMARY

Some valuable lessons were learnt through the formative and summative evaluations. It was assumed that the Alfresco system would be user-friendly for the members of both case studies, but the answers received from respondents showed that they did not feel the system was intuitive enough. The system was found to be sufficient with regards to the backing-up of data, and workflow management in a closed environment, but it was not fully integrated with daily tools and work processes (in the laboratory or elsewhere), which could be the reason why its usage was limited.

The lack of usage of metadata schemas, or other metadata elements, such as tags and categories, are concerning, especially when the members in these groups later reach the stage where they want to publish their data in a repository or elsewhere. A more thorough awareness campaign will have to be launched to inform researchers about the value and necessity of metadata.

The training sessions presented by the VRE-D and the researcher of this study were found to be insufficient and should have been followed up with more training sessions to ensure that everyone was on the same page.

The responses received from the librarian showed that there is a lack of awareness among librarians about the valuable role they can play in VREs. Bigger effort should be made to train, inform and up-skill librarians about VREs and RDM, so that they can take their place as fully-fledged members in these VREs, providing the necessary consultancy, guidance and training to researchers. Librarians, in addition to information / data searches can, for example, provide guidance on metadata schemas, referencing of data, file naming conventions, publishing of data, licensing of data, and copyright, etc.

The results also showed that protocols as formal documents that guide actions within the VRE, could be added as a possible policy component in the conceptual VRE model discussed in 5.4. An effort should also be made to raise awareness among researchers about the value of DMPs and its linkage to the ethical processes at the University. Training should also be given to researchers on how to compile a data DMP, which is something that could typically be performed by librarians.

Although not specifically asked, none of the respondents mentioned funder requirements that data should be published on a data repository, and that researchers should have a DMP. Funders were listed in 3.5.7.1 as a potential human component of a VRE. The reason this was not mentioned could be because the respondents, at the time of the interviews, had not reached the stage where they had to publish their data.

The Research Office, which was also mentioned as a potential human component in 3.5.7.1, plays an important role in ensuring that researchers comply with funders' requirements. This was also not mentioned by the respondents, but the office would need to be involved once the researchers reached the stage where they publish their results and data.

The next chapter includes a discussion on how the findings in the literature as well as empirical study answer the research question and sub-questions. This is followed by a reflection on the study, an overview of the contribution of this study to the subject field, a discussion of the limitations of the study, some recommendations, suggestions for further study, and concluding remarks.

CHAPTER 8
CONCLUSIONS AND RECOMMENDATIONS

8.1 INTRODUCTION

In Chapter 1 it was mentioned that the VREs that have been built thus far have tended to be either precise configurations for specific research projects, or systems having very generic functions. These 'systems', as pointed out by Voss and Procter (2009: 176), have had "significant fragmentation" and a shortage of interoperability, which necessitated "agreed standard platforms and configurable modules" that will enable swift development and implementation of tailored VREs. The researcher of this study then identified a need for the formalisation of a conceptual model of a VRE that could be used repeatedly in different contexts and different subject fields. In addition, the researcher decided to investigate what the relationship between VREs and RDM is, and whether or not a VRE should be an essential framework for the management of research data. From this flowed the central research question and its sub-questions. The aim of this chapter is to address these questions from the findings in the empirical part of the study, corroborate these from findings in the literature study, and to draw conclusions from these. This is followed by a reflection on the findings from the case studies and literature, a discussion about the contribution of this study to the subject field, and an indication of the limitations of this study, guidelines and recommendations for setting up a conceptual model, suggestions for further research, and concluding remarks.

8.2 CENTRAL RESEARCH QUESTION

The central research question, as stated in 1.2 was: How can a Virtual Research Environment be conceptualised to indicate the role of Research Data Management (RDM) within a VRE?

To answer this question a number of sub-questions were asked. These are:

- What is a VRE?
- What is the current state of VRE research in the world?
- What are the generic components that make up a VRE?

- How does a VRE support a research cycle?
- What is RDM?
- Why should a VRE be an essential technological and collaborative framework for the management of research data?
- To what extent can the components identified in the third sub-question be formalised into a conceptual framework?
- Where would RDM as component be placed?
- To what extent can this model be generalised for use in other environments?
- How was the central research question answered?

The answers to these questions are discussed next.

8.2.1 What Is A VRE?

The literature study revealed that there are a number of concepts that are closely related to the concept of a VRE. These are e-Science, cyberinfrastructure, science gateways, cyberscience, e-Research, collaboratories, and WRSS. Although the respondents in the empirical study mentioned none of these concepts, they did mention some of the characteristics of these concepts.

The discussion in 2.3 showed that VREs contribute to the broadening of the definition of e-Science from grid-based distributed computing for scientists with huge amounts of data, to a definition that includes the development of online tools, content, and middleware within a coherent framework for all disciplines and all types of research. This is in line with the results of the empirical study. Case Study A had been using natural science-oriented data, and laboratory/experimental methods, and had been using a VRE to upload these, whereas Case Study B had been using human-oriented data and survey instruments as data collection method, and had been using a VRE to upload these (see 7.3.1.1, Question 3). The answers received from the respondents in Question 3 of 7.3.1.1 also confirmed that there was a need for multi-disciplinarity in VREs. The discussion in 2.3 further mentioned that 'cyberinfrastructure' refers to all the aspects of the digital side of research infrastructure, with VREs as the interface to that infrastructure. Science gateways were also described in 2.3 of the literature study, as convenient interfaces to cyberinfrastructure, which showed that VREs and science gateways are synonymous. This aspect of providing an interface to cyberinfrastructure corresponds to one of the future objectives of a VRE as mentioned in the empirical study under 7.3.1.1, Question 21 (c)(vii), namely the establishment of an interface similar to that of Facebook, which would be more intuitive.

In 2.2.7, the discussion on WRSS revealed that these types of web-based systems are used to support research institutions and researchers, in the finding of relevant information. This links up with the idea mentioned in 7.3.1.1 under Question 12, that literature searches and search strategies should be included as part of Research Data Management. It also links up with the answer received in 7.3.1.1 under Question 15 (a)(v), that searches and search strategies should be part of tasks and actions performed in a VRE. WRSS were also shown in 2.2.7 as systems that are used in the development of new and effective tools, in choosing the right tools for research, and in improving the quality of the presentation of research results. The aspect of affording researchers the opportunity to develop their own tools was confirmed by the VRE-D in the empirical part of this study in 7.3.1.3 under Question 44. The discussion in 2.3 further showed that web-based VREs are synonymous with WRSS, because they both contribute to research support systems and provide collaborative work support.

The concept of collaboratories was shown in 2.3 to be totally synonymous with that of VREs and that it appeared to have been supplanted by the VRE concept. This could maybe be the reason why none of the respondents in the empirical study mentioned the concept.

Another concept closely related to VREs was shown to be the concept of e-Research. In 2.2.5, e-Research was described as a broad term that extends to e-Science, and also as a form of scholarship that is conducted in a networked environment that encompasses all information and communication technologies (ICTs) that support researchers in their research process. This incorporates all forms of non-computational e-Science, consisting of a wide array of new technologies, tools and computer networks, which can be used collaboratively by researchers that are co-located or separated by distance globally. E-Research was used as the framework from which the central research question was investigated.

The literature study in 2.3 indicated that there are various approaches to e-Research, and that this study followed the social sciences approach to e-Research, which included the computerisation movement, information systems, SOA, and whole process approaches. The computerisation movement approach was revealed as an approach that focuses on computer-based systems as instruments to bring about a new social transformation, and advances the development of new information infrastructures with their accompanying technologies for research, as well as the application of these in varied ways across research fields, disciplines and scientific institutions. It also provided a valuable framework that was used to understand and explore the application of these technologies in VREs, and their effect on organisational transformation. The use of VREs in the two case studies (empirical study) it would seem, brought about some organisational transformation. In 7.3.1.1, under Question 5, it was revealed that VREs made it possible for members to access data from anywhere and to share information, files and data with one another from any geographical location in the world. The VREs also provided the ability to interact and collaborate with members across any geographical distance (see 7.3.1.1, Question 5). The interaction and collaboration between members in Case Study A, however, were shown to be limited, because the social aspect of their VRE had not been developed much yet, whereas the social aspect in Case Study B had been developing faster (see 7.3.1.1, Question 8). The Workflow function of the VREs was also used extensively by the members of Case Study B and had a direct impact on the way members interacted and shared information or data. This was confirmed by comments received in 7.3.1.1, under Question 15 (h).

The discussion on the information systems approach in 2.2.5.1 revealed that this approach clarifies and formalises domains of human activity, and creates interventions by IT-based systems in those domains. VREs with their interoperating range of online tools, network resources and technologies were also shown to have an impact on researchers and their research. The empirical study in 7.3.1.1 under Question 19, however, revealed that there had been no formal policy in place in either case study. In Case Study A, there was a mutually agreed project plan and collaborative agreement on how data should be uploaded onto and archived in the VRE, and shared through the VRE. In Case Study B, protocols as formal documents guided the members’ actions within the VRE.

The SOA approach to e-Research, which was discussed in 2.2.5.1, also divulged that this approach could be used to ensure that VREs are flexible enough for dynamic user needs. This corresponds to the researcher of this study's definition of a VRE in 2.2.8.1, where he defined a VRE as a "common, flexible, technological and collaborative framework." The various cycles of formative evaluation as given in the formative evaluation part of the empirical study, revealed further that the VREs of both case studies (see 7.2.1 and 7.2.2) were adapted (in other words were flexible) to fit users' needs.

The whole-process approach as discussed in 2.2.5.1, was adopted because it included the development of demonstrator models (prototypes) "to illustrate how the process will work in practice" (Paterson, 2007: 128). During the formative evaluation part of the empirical study in 7.2.1.4, the initial VRE prototype was demonstrated, using Moodle software that was linked to a DSpace instance. This was also confirmed by the VRE-D in the summative evaluation part of the empirical study in 7.3.1.3, where he indicated that he had used Moodle to develop a prototype of a VRE.

The literature study showed in 2.2.8.1 that there are a wide variety of definitions of VREs. The core elements of these definitions were subsequently consolidated by the researcher of this study in the following definition:

> A VRE consists of a common, flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts etc.) interoperating with each other, can be plugged, to enable collaboration and to support and enhance large and small scale processes of research, which are often performed by researchers in multidisciplinary contexts within or across organisational and geographical boundaries.

All the elements of this definition of a VRE, as well as the characteristics of a VRE listed in 2.2.8.3, were confirmed by answers received from respondents in the empirical study in 7.3.1.1 under Question 5 and also in Table 7.1. In 2.2.8.3, it was also mentioned that a VRE is typically project-driven. This was confirmed in 7.3.1.1, Question 4, where the majority of respondents of both case studies were of the opinion that the VREs were driven by the topics of the VREs, and that their components had been designed around the needs in the various projects within them.

Another characteristic mentioned in 2.2.8.2, was that VREs are designed strategically rather than responsively or incrementally. This was verified by the answers received from the VRE Managers in 7.3.1.2, Question 24, as well as in the results from the formative evaluation in 7.2.1.1, 7.2.1.2, 7.2.1.3, 7.2.1.4 and 7.2.2.1. Meetings were held with each of the promotors (later called VRE Managers) of these groups, which were followed up with meetings with the student researchers and the VRE Champion for Case Study A, together with the promotor of Case Study A, in order to clarify issues such as confidentiality and securing their data, and to discuss their needs and requirements. In a similar manner, a meeting was held with the promotor of Case Study B and her student researchers, to clarify their needs and requirements with regards to a VRE.

Yet another characteristic mentioned in 2.2.8.3, is that VREs facilitate collaboration amongst researchers and research teams, providing them with more effective means of collaboratively collecting, manipulating and managing data, as well as collaborative knowledge creation. This was corroborated by the answers received from respondents in both case studies in the empirical study. In 7.3.1.1, Question 8, it was mentioned that a VRE provides a central platform or database to share information (articles, documents and data). The 'Shared Files' component also encouraged collaboration amongst members of these groups around shared files and shared folders. Members could use these files and create and update them. The 'My Tasks' workflow component in 7.3.1.1, Question 8, further afforded members of Case Study B the ability to work on files collaboratively.

In 2.2.8.3, it was mentioned that a VRE normally has a web-based front end (or portal), which enables researchers to access the VRE via a web browser using a personal computer or mobile devices such as cell phones and tablets. This was validated by the empirical study in 7.3.1.1, Question 21 (c) (vii), where one of the future objectives mentioned by one of the respondents (A-R5) was having a VRE with an interface similar to that of Facebook. It was also mentioned in 7.2.2.5 that the interface of Alfresco could be accessed through a web URL, http://icarus.up.ac.za:8080/share, using a desktop computer and logging into the system using a username and password. In 7.3.1.1, Question 8, it was further disclosed that files could be synchronised through an app using a mobile device, or could be synchronised using a desktop computer.

Yet another characteristic of a VRE mentioned in 2.2.8.3 is that it can be described as a one-stop shop where researchers can obtain data and global information pertinent to their research with suitable “semantic support and contextual services for discovery, location, and digital rights management” (Yang and Allan, 2010: 68). This was substantiated by findings from the empirical study. The empirical study in 7.3.1.1 under Question 8 disclosed that the Alfresco VREs used in the two case studies, had a ‘search’ component that had been used by some of the respondents to discover and locate files, documents and people in their VRE.

In the empirical study under 7.2.1.7, it was revealed that each of the members in the two case studies had rights to access and edit their own spaces and to read and access shared spaces in the VRE, while the supervisors/promotors had VRE Manager rights and the Laboratory Manager acting as VRE Champion also had VRE Manager rights, which gave them rights to access all the members’ spaces in their respective VREs. These rights were set under the ‘Manage Permissions’ component at 7.3.1.1, Question 8. In 7.3.1.1, Question 16 (c), it was also mentioned that the terms of use of a dataset could be described by adding a copyright license, for example a Creative Commons License, to the data. This could be added as a field in the metadata record.

In 2.2.8.3 it was stated that a VRE could be constructed on top of existing applications such as VLEs. The empirical study in 7.2.1.4; 7.3.1.2, Question 24; and 7.3.1.3, Question 34, confirmed that it is possible to construct a VRE on top of a VLE. It was shown that Moodle, a VLE platform, was used to create the first prototype of a VRE for Case Study A. Moodle, however, proved to be insufficient for the needs of the respondents of the two case studies and was later replaced by Alfresco, a document management system.

The discussion on characteristics in 2.2.8.3 described VREs as the products of “joining together new and existing components in support of as much of the research process” as possible for any activity (Fraser, 2005; Wilson, et al., 2007: 290). This was corroborated by findings in the empirical study. In the formative evaluation, part of the empirical study under 7.2.5, those functionalities and components that were seen as important, and that were well-used by individuals in these groups, were identified, for example, archiving and back-up of data, versioning, the workflow function, e-mail notifications of actions happening on the VRE, and synchronization of files on a desktop computer via an application (app) to the VRE. A range of new functionalities were also pinpointed, such as affixing a link from the VRE to Google Drive, a survey tool, a supplementary storage device for Case Study A, the replication of this storage device to one in the library, the backing-up of devices and apparatus in the laboratory and the connection of these to the network, and the provision of access to a virtual machine environment to one of the student researchers to test some of the software and processing power.

The literature review in 2.2.8.3 revealed that VREs could be used for analysis and processing of data, annotation of data collaboratively, and sharing of data with peers. The empirical study, however, showed in 7.3.1.1, Question 15 (e) and (f), that the Alfresco platform used for these two case studies, could not be used for analysis and processing of data. Members then expressed their need for this component to be added and integrated into the VRE platform. The VRE-D revealed that it was already available within the HUBzero platform, which he was investigating as a potential future VRE platform. In 5.4 and 4.5.2.3, data analysis tools were also mentioned as RDM tools that could be added to a VRE. The aspect of sharing of data through a VRE was confirmed in the empirical study in 7.3.1.1, Questions 5 and 8, while the aspect of annotating data was done through the ‘Comments’ component mentioned in 7.3.1.1, Question 8. The results showed that members of Case Study B had used the ‘Comments’ function more than members from Case Study A.

The discussion in 2.2.8.3 revealed that VREs enable inter-disciplinarity, by bringing data and approaches from different disciplines together to create new research findings. This was validated by findings in the empirical study in 7.3.1.1, Question 3, where most of the respondents in Case Study A identified their VRE as catering for more than one discipline. One respondent even described it as one discipline cutting across other disciplines or fields. In Case Study B, there were also two respondents that were involved in multi-disciplinary projects.

In 2.2.8.3, it was mentioned that a VRE can be technology-driven, but preferably demand-driven, which will ensure that they are end-user focused. This was confirmed by responses received in the empirical study in 7.3.1.1 under Question 4, where three of the respondents in Case Study A were of the opinion that their VRE was technology driven, and four were of the view that their VRE was driven by the topic of their research. In Case Study B, all the respondents were of the view that their VRE was driven by the topics of their research; in other words, the VRE and its components had been designed around the needs found in the various projects.

Another characteristic mentioned in 2.2.8.3, was that a VRE system should be able to act as a communication platform. This was proven to be correct through the empirical study in 7.3.1.1, Question 5, where it was mentioned by A-R5 that a VRE could be used to communicate by using information technology. The B-L also stated in 7.3.1.4, Question 45, that the VRE provided a platform for her to communicate with researchers.

The literature study in 2.2.8.3 disclosed that VRE systems should be as flexible as possible because user requirements are constantly changing. This was proven, as mentioned earlier, by the various cycles of formative evaluation that took place in the formative evaluation part of the empirical study. This revealed that the VREs of both case studies (see 7.2.1 and 7.2.2.) were adapted (in other words were flexible) to fit users' needs.

In 2.2.8.3, it was mentioned that VREs can follow a three-tier or multi-tier (n-tier) architecture, where web portals can act as the presentation layer, with business logic and data layers behind it. The conceptual framework model that was proposed in 5.4, illustrated this multi-tiered architecture in more detail. The model consists of a human layer with possible human components as a first tier, a hardware layer with possible hardware components as a second tier, and a software layer, comprising possible software components, as a third tier. In addition, the proposed VRE has policy components, a management services component, and a standards, specifications and protocols component. For a more detailed discussion of the different layers and components, see 8.2.3.

8.2.2 What Is The Current State Of VRE Research In The World?

The researcher conducted a literature study on the current state of VRE research in the world, in 3.1. Four countries were selected as representative of different VRE approaches or models used across the globe, namely the UK, the USA, the Netherlands, and Germany. It was mentioned that VRE programmes in these countries, though each unique in their own way, share a relatively similar vision of key elements of VREs, and they are specifically aimed at facilitating the shared use of digital infrastructure by researchers through the provision of shared environments. In 3.3, the researcher discussed the similarities and differences in the VRE programmes of these different countries by looking at organisational, technical, functional, policy / legal / financial, and cultural aspects.

Under the organisational aspects in 3.3.1, it was revealed that the German DFG required its funded projects to be collaborations between researchers and infrastructure developing institutions, such as libraries, computer centres, and e-Research centres. The empirical study in 7.2.1 and 7.2.2 further revealed that, although not funded projects, Case Study A and Case Study B were collaborations between researchers and the Department of Library Services as an infrastructure developing institution at UP. The discussion in 3.3.1 further revealed the importance that was placed in the UK on including the users of these projects in the design process. JISC's use of a Figure 8 Participative design process, where users and developers design a VRE together, was also mentioned. This process included a user needs analysis as well as a contextual and change analysis among the users, an analysis and design of systems, and the building of VRE pilots while keeping quality assurance in mind. The literature study also showed a similar bottom-up and user-driven approach in the USA with regards to the technology and software used. Similarly, the SURFNet programme in the Netherlands and the DFG programmes in Germany gave users the freedom to experiment and to develop their own technologies, or adapt existing ones. The empirical study showed in 7.2.1.3 that a similar approach as in the UK, USA and the Netherlands was followed. This included a bottom-up approach where members of Case Study A could collaboratively give their inputs in the design / setup of the VRE. This process consisted of contact sessions by the VRE-D and the researcher of this study, with members of Case Study A, where a needs analysis was done (see 7.2.1.3, 7.2.1.5, 7.2.1.7, 7.2.1.10, 7.2.1.13 and 7.2.1.14). In the same manner, a needs analysis of users in Case Study B was done through a number of contact sessions (see 7.2.2.3, 7.2.2.4, 7.2.2.7, 7.2.2.8, 7.2.2.11, 7.2.2.12, 7.2.2.14, 7.2.2.15, 7.2.2.17, 7.2.2.18, 7.2.2.20, and 7.2.2.21). VRE pilots were developed collaboratively for both case studies (see 7.3.1.2, Question 24, and 7.2.1.1).

The discussion in 3.1.2 under the technical aspects disclosed that some of the UK VRE projects had made use of shelf-ready tools to create VREs, for example Sakai and Moodle, which are actually VLE tools. Others used content/document management tools, for example SharePoint, while some used portal technologies, and others used general institutional web-based tools. In contrast, the German DFG encouraged and funded projects that developed new software using open source principles, but also funded projects that applied existing solutions. An overview of developments in the USA revealed the development of science gateways (or portal technology and gridware), as well as the creation of hubs (cloud driven tools), while a study of developments in the Netherlands disclosed a flexible approach, where funded projects were given the freedom to test any environment that would meet their needs. The empirical study in 7.2.1.4 revealed that the first prototype for Case Study A was developed using Moodle, a VLE platform, which was very much in line with developments in the UK. This was followed-up by a development of a VRE for Case Study A (see 7.2.1.8) and Case Study B (see 7.2.2.4) using Alfresco, a document management system, which was also in line with the developments in the UK.

The functional aspects discussed in 3.3.3 revealed that the VRE programmes in all four countries focused on collaboration and sharing, supporting the research lifecycle, and on supporting single, interdisciplinary and cross-institutional research. The literature on Dutch and German projects revealed that data sharing was an important and central feature, whereas the US Science Gateways project also disclosed that there was a growing trend to make data available. The UK programme, on the other hand, emphasised the importance of evaluating/assessing the success of VRE projects, while the Science Gateways project in the USA discovered that VREs were convenient interfaces to supercomputing resources. The sharing of data was confirmed by the empirical study, which showed that one of the important characteristics of Alfresco was that it enabled users to do sharing of files (including data) (see 7.2.1.8 (b); 7.3.1.1, Question 5; 7.3.1.1, Question 8; 7.3.1.1, Question 13; 7.3.1.1, Question 21). Respondents also used the ‘Shared Files’ and ‘Share Data with Peers’ components for the sharing of data (see 7.3.1.1, Question 8). The idea of using VREs in collaboration and sharing of data in interdisciplinary research and cross cross-institutional research, was confirmed in 7.3.1.1, Questions 5 and 13, where it was mentioned that VREs can stretch across organisational and/or geographical boundaries. The empirical study also disclosed that there was a need for VREs to support interdisciplinary research (see 7.3.1.1, Question 21 (c) (iv) and 7.3.1.1, Question 3). The finding in the Science Gateways project in the USA that VREs were interfaces to supercomputing resources, was validated by a need expressed by the respondents in the empirical study, namely the addition of processing and computing power to the VRE, and the establishment of a high-performance-computing set-up (see 7.3.1.1, Question 21 (c)(iv) and (viii)). In 7.2.1.4, it was further mentioned that the folders in the original prototype were structured in such a manner that it would support the research lifecycle, something that was mentioned in all four countries’ VRE programmes.

The discussion of policy / legal / financial aspects in 3.3.4 showed that three of the countries - the UK, the Netherlands and Germany, each had a national institution that funded and drove the major VRE initiatives in their respective countries. This is different, however, in the USA, where only TeraGrid, a sub-project of the Science Gateways project, was funded by the National Science Foundation. In addition, the UK has had a joint government-commercial venture between the British Library and Microsoft from 2007-2013, to develop the RIC project (Research Information Centre Framework, 2016). The discussion in 3.3.4 further revealed that a major challenge faced by all the VREs was sustainability, especially with regards to long-term funding, development of business models to make VREs self-sustaining, and the acceptance and use by communities they were aimed at. The provision of funding for infrastructure for the management of research data was mentioned by one of the respondents (A-R4) as an aspect that should be included in a university RDM policy (see 7.3.1.1, Question 21 (c)(vi)). Other issues that were raised by Dutch SURFshare, the German eSciDoc project, and the UK myExperiment were more applicable to the librarian’s domain. The first issue dealt with the sharing of resources that require institutional subscriptions, with researchers at other institutions that do not have subscriptions (Dutch SURFshare). None of the respondents in the empirical study mentioned this issue, however. This included the librarian (B-L), although this could be a typical issue where the librarian could add some value. The reason the B-L did not mention this, could be because she had not been certain of the potential role she could play in the VRE (see 7.3.1.4, Question 47). The second issue dealt with, was the importance of having librarians that could check before publication, that third party rights had not been violated by open access publications. The third issue dealt with the aspect that the sharing of all data (total access to everything) was not always possible, but that some data might have certain usage or access rights. This role of checking that third party rights had not been violated, together with giving advice on copyright and licensing of data, which determined access to and usage of data, were identified as potential roles a librarian could play (see 7.4), but were not mentioned by the B-L in 7.3.1.4, Question 47. The reason this were not mentioned could be because the respondents in Case Study B had at the time of the interviews not yet published their data.

The cultural aspects discussed in 3.1.1 disclosed that the UK was well advanced in its understanding of the VRE concept and had the world's best-structured programme of VRE developments so far. Projects in the Netherlands revealed more of a focus on the humanities and social sciences, while focus in other countries was shown to be more multi-disciplinary. The German projects, on the other hand, revealed that the building of appropriate services and solutions that facilitate collaboration across discipline boundaries, were complicated. The empirical study in 7.2.1 and 7.2.2 showed that VREs could be developed for different disciplines using the same software platform (in this case Alfresco), namely Case Study A for natural science-oriented data and laboratory / experimental methods, and Case Study B for human-oriented data, using survey instruments as data collection method. Some of the respondents in 7.3.1.1, Question 3, also revealed that some of their projects were interdisciplinary in nature. The German project eSciDoc highlighted trust as a key factor in the uptake of a VRE. This issue of trust also came to the fore during the empirical study in 7.3.1.1, Question 8, where one of the respondents, A-R4, revealed that he didn't trust the system to keep all his versions, and he expressed his anxiety that his files might get corrupted. A-R 4 also revealed in 7.3.1.1, Question 15 (b) that he had used additional tools for storage and backup of his data, signalling a motion of distrust in the VRE as sole tool for storage and back-up. The discussion in 3.1.1 furthermore disclosed that VRE projects in all four countries focused on supporting the research lifecycle. None of the respondents in the empirical study, however, mentioned that VREs should focus on supporting the research lifecycle. Answers received during the interviews, nevertheless, revealed that various stages of the research lifecycle and research data lifecycle were supported through the VRE (see discussion in 8.2.4). The B-VRE-M and the VRE-D, on the other hand, did mention RDM in terms of the research data lifecycle.

8.2.3 What Are The Generic Components That Make Up A VRE?

In 8.2.1, the characteristic of VREs having a multi-tier architecture, were discussed. This discussion revealed that the potential components of a VRE can be grouped in a human layer with possible human components, a hardware layer with possible hardware components, and a software layer, comprising possible software components that interact with each other. In addition, it was mentioned that a VRE had vertical component layers, namely a policy component, a management services component, and a standards, specifications and protocols component.

The literature study in 3.5.7 and 5.4 revealed that the human components layer formed the first layer, and could consist of a core group, for example, researchers, VRE Managers, VRE Champions, Librarians, Research Office, University IT, University Executive, and a peripheral group, for example developer(s) (designers), funders, peer reviewers, the community, and publishers. In 6.2.2.1, it was disclosed that the human components layer in Case Study A consisted of the following components: student researchers, a VRE Manager, and a VRE Champion, while Case Study B had a VRE Manager, student researchers and a librarian. Both case studies shared a VRE Designer (see 6.2.2.1). In 7.2.2.1, the B-VRE-M also mentioned that they would need to provide publishers with the possibility to interrogate their data. The empirical study further revealed in 7.2.1.1 that the Chair of the Ethics Committee was included in a meeting on 11 April 2013 to discuss the possibility of a VRE pilot study in the Faculty where Case Study A was situated. The Ethics Committee, as a human component, was not found in the literature study, however, but could be added to the core group of human components.

The literature study in 3.5.7 and 5.4 disclosed that the second tier of a VRE was a hardware component layer. This, according to 3.5.7, was made up of four categories:

desktop services, e.g. personal computers (PCs); mobile devices, e.g. laptop computers, notebook computers, netbooks, computer tablets, or cell phones; data capture and output devices, e.g. digital still cameras, digital video cameras, and digital recorders such as digital pens and voice recorders; as well as cyberinfrastructure, including local networks (e.g. servers), the national backbone, and international infrastructure (e.g. cloud services). The empirical study 7.3.1.3, Question 39, revealed that the hardware component of the two case studies consisted of three servers, which all replicated each other. All three of these servers were linked to the University's backbone/network. Other hardware components used by respondents, as revealed in the empirical study, included desktop and laptop computers (see 7.3.1.1, Question 8, under 'Desktop Syncing'). The empirical study further revealed in 7.2.1.8 (b) and 7.2.2.6 that a mobile app was available for download onto a mobile device, but only one member of Case Study A, and only two members of Case Study B used this app via their mobile phones to synchronise with the Alfresco VRE system (see 7.3.1.1, Question 8). The empirical study also disclosed in 7.3.1.1, Question 15 (a)(i), that members in Case Study B had been using Apple iPads (tablets) to do video and sound recordings, which were then uploaded onto the VRE. In addition, respondents in both case studies revealed that there existed a need to plug-in their laboratory and other instruments to the VRE, so that they could access it through the VRE, and so that they could upload / capture data automatically into the VRE, from instruments. Furthermore, in 7.3.1.1, Question 17, one of the respondents also mentioned the use of a machine that measures the inner ear in conjunction with Video Nystagmography that measures different eye movements. This machine could also potentially be plugged into the VRE, in order to enable the upload of data generated through it, directly into the VRE.

The third layer of the VRE as identified in 5.4 and Figure 5.2b was a software components layer, consisting of an interface/platform in the form of a web portal, VLE, or a proprietary tool. This was confirmed in 7.2.1.5 (b) and in 7.3.1.2 under Question 24, where it was mentioned that the first VRE site created for Case Study A was done by adapting Moodle, a VLE system, as a VRE platform. The Moodle platform was subsequently replaced by adapting Alfresco, a document management system, as a VRE platform for Case Study A (see 7.2.1.8 (b)). The idea of the VRE being a platform was confirmed by one of the respondents that mentioned that a VRE is "something that a person is connected to online … [the person] is [then] able to share the information on one's central platform" (see 7.3.1.1, under Question 5). The VRE-D also confirmed that he had experimented with a number of open source tools that could be adapted as VRE platforms, for example Moodle, Chisimba, and Alfresco. At the time of the interview, he was also experimenting with an open source platform, called HUBzero, which was specifically designed as a VRE platform by Purdue University (see 7.3.1.3, Question 35).

In 3.5.7.3, it was pointed out that as part of the interface or platform, there is also an authentication layer to determine the level of access a human component can have to the software layer. This was confirmed by the A-VRE-C in 7.3.1.2, Question 25, where she indicated that she controlled the authentication and levels of permissions in the group. She created a user profile for each new researcher student on the system, and invited them to join the group on the VRE. She also disabled their access when they left. In 7.2.1.5, it was confirmed that Alfresco used a password authentication, giving access through a username and password. This corresponded to one of the authentication methods mentioned in 3.5.7.3. The VRE-D in 7.3.1.3, under Question 44, also mentioned the possibility of replacing Alfresco in the near future with HUBzero and adding this to the University's authentication directory, for example the active directory, so that any researcher can sign onto the VRE.

The software layer as indicated in 3.5.7.3 also includes a core interface/software layer consisting of fixed components that are part of the standard configuration of the specific tool used. These fixed components could vary, but are normally things such as a search function; a personal profile; collaborative writing tools such as blogs and wikis; communication tools such as instant messaging, chat, and e-mail; a document store where researchers are able to create new versions of documents and store them for publishing at a later stage, if they so wish (a document management function); a RDM component, e.g. a research data store (more components could be added to enhance the RDM functionality); a settings function; a site news function; a site admin function; and a calendar. Sometimes it can also include some of the components that have been listed in the pluggable components layer. The empirical study in 7.3.1.1, Question 8, and 7.3.2, disclosed that the Alfresco VRE had the following core components, although not all these components were used by the respondents in the case studies: 'Search', 'Edit your profile', 'Site Calendar', 'My discussions', 'Following', 'My Files', 'My Activities', ‘Site Activities’, ‘My Tasks’ (Workflow Function), ‘My Documents’, ‘Shared Files’, ‘People Finder’, ‘Invite Users’, ‘Discussions’, ‘Document Library’, ‘Publishing Function’ (blogs and wikis).

The fourth layer of a VRE as mentioned in 5.4, consists of RDM components. A number of these RDM components were confirmed by the empirical study. Data capturing tools were mentioned in 7.3.1.1, Question 15 (a) (i), while data processing tools were mentioned in 7.3.1.1, Question 15 (e). The empirical study in 7.3.1.1, Question 8, further revealed that there was a publishing function in the Alfresco platform, but that this only published to social media. A need to plug-in a repository tool where data could be published, were mentioned, although the groups had not yet reached the publishing stage at the time of the interviews (see 7.3.1.1, Question 15 (j); and 7.3.1.2, Question 29). Data management planning tools were not mentioned by any of the respondents as possible tools to add to the VRE, but the importance of DMPs was mentioned by the VRE-D in 7.3.1.1 under Question 12. The VRE D also mentioned in 7.3.1.1, Question 15 (j), the importance of plugging preservation tools, for example BagIt, into the Alfresco VRE. The empirical study in 7.2.1.8 (b) and (c) further disclosed that the Alfresco platform had a very good metadata function (Dublin Core Metadata template) built into its core. Unfortunately, none of the respondents in either of the case studies had made use of the Dublin Core Metadata template (see Question 8), and none of respondents had mentioned the possibility of adding a metadata store to the VRE, which showed a lack of understanding of its importance.

The aspect of providing access to computational resources was validated by the A-VRE-M in 7.3.1.3, Question 21 (c) (iv), when he identified that their VRE would need more computing power in the future. The need for data analysis tools/software to be linked to the VRE platform was something that was mentioned by B-R3 in 7.3.1.1, Question 8, as well as B-VRE-M in 7.3.1.1, Question 15 (f). It was also a need that was expressed by members of Case Study A in 7.2.1.5. The visualisation of data through the VRE was another need identified in 7.3.1.1, Question 15 (g). Some of the researchers (A-R3, A-R4 and A-VRE-C) mentioned that they had been generating visualisations of their data, and A-R4 indicated that it would be valuable to have a visualisation tool available within the VRE. In Case Study B, the B-VRE-M communicated that the projects within group had reached the stage where they would need to be able to generate visualisations within the VRE (see 7.3.1.1, Question 15 (g)). The VRE-D confirmed that data visualisation could not be done within the Alfresco platform, which meant that a data visualisation tool would need to be plugged-in. He did, however, mention that it is part of the core functions within the HUBzero platform.

The discussions in 3.5.7.3 and 5.4 identified a bottom layer in the software components layer that comprises various software components that could be plugged into the interface/platform component, and are determined by the needs of each VRE community/project. A number of possible components were listed, which were confirmed by the empirical study. Document management tools were identified in 7.3.1.1, Question 34 as tools that are essential in terms of managing one's data. It was pinpointed as some of the major reasons why the instance of the VRE on the Moodle platform was replaced with an instance on Alfresco. The Alfresco platform already had a document management function built into it as part of the core function. Another pluggable component that was mentioned in 3.5.7.3 was specialist computational software that would, for example, enable the usage of HPC and sequencing. The empirical study confirmed this in 7.3.1.1, Question 21 (c)(iv), where it was stated that the A-R3, A-VRE-M and A-VRE-C had mentioned that the VRE platform would need processing power, computing power, and more storage space. The A-VRE-M also elaborated in 7.3.1.1, Question 21 (c)(viii) on the plans for the future to establish a high-performance computing set-up for Case Study A.

The empirical study in 7.3.1.1, Question 21 (c)(ii), also revealed that the VRE instance on Alfresco did not have sequencing tools and simulation tools built into it, and that there was a need for sequencing tools and simulation tools to be plugged into the system. E-learning tools and skills development tools were identified in the literature study in 3.5.7.3 as pluggable components, but none of the respondents in the empirical study indicated a need for this. In the literature study, modelling tools were also mentioned, but in the empirical study, in 3.5.7.3, the respondents in both case studies failed to mention this. In 5.3, modelling tools were shown to be used synonymously with simulation tools by Martinez-Uribe and Macdonald (2009: 311), which confirmed the inclusion of these in the software pluggable components layer. Geospatial tools were mentioned next in the literature study, as a possible pluggable component, and in the empirical study in 7.3.1.1, Question 15 (d), the VRE-D pointed out that the Alfresco system pulled metadata automatically from the system, including geographical location (GPS coordinates). In 7.3.1.3, Question 38, the VRE-D further indicated that he had enabled the Google Maps integration with Alfresco, which showed where photos were taken.

None of the respondents mentioned intellectual property management tools, although the issue of intellectual property was mentioned by die A-VRE-C in 7.3.1.1, Question 16 (b). Access to electronic information sources was mentioned in 3.5.7.3 as a potential component, but none of the respondents in the empirical study mentioned this as something that they would want to be integrated with the VRE. This could be because they had not considered the option of searching for information sources within the VRE. Referencing tools were identified in 3.5.7.3 as potential pluggable tools. This was then validated in 7.3.1.1 under Question 9, where the need was mentioned for the VRE system to link to a referencing system (tool), for example EndNote or Refworks. A DOI generator was mentioned in the literature as a possible plug-in, but none of the respondents in either of the case studies identified this as a possible component. The reason for this could be because none of the respondents had published data yet at the time the interviews were held.

Experimentation tools were mentioned in 3.5.7.3 in the literature study as another pluggable tool. In 7.3.1.1, Question 15 (a)(iii), one of the respondents (A-R1) in Case Study A indicated that experiments could be done through, for example, bioinformatics and programming experiments via a computer. To be able to do it within the Alfresco VRE, however, one would have to get the rights from the people who developed the experimental software. The VRE-D stressed that it would be possible to do experiments within the HUBzero platform (it is part of its core). As mentioned earlier, he plans to migrate the two case study VREs from Alfresco to HUBzero.

Another component that was mentioned in 3.5.7.3 was access to remote instrumentation. This was confirmed in the empirical study in 7.2.1.14 (b), where the VRE-D provided one of the student researchers (A-R2) access to a virtual machine environment. Electronic lab books were introduced in 3.5.7.3 as possible pluggable tools, but the empirical study revealed in 7.2.1.2 that the group preferred to keep their lab books in paper format during their projects, and then digitise these at the end of the study. The reason for this has probably to do with entrenched work practices, and legal and ethical issues, where researchers need to be able to provide activities in written form in a lab book, as proof of the work they had done. It could also indicate a lack of trust in the legality of work done on electronic lab books.

In 2.2.8.3, it was stated that a VRE should have the following three components: a recording process (capturing data), clear ownership (through authentication) of the data, and a focus on a specific question or topic. The recording process/data capturing process was touched on by respondents in the empirical study in 7.3.1.1, Question 15 (a)(i). In the empirical study in 7.3.1.2, Question 25, the A-VRE-C revealed that she controlled the authentication and levels of permissions for members of the Alfresco VRE in Case Study A. The last of these three components were confirmed in the empirical study in 7.2.1.7 and 7.3.1.1, Question 4, as well as in the summary of the summative evaluation in 7.3.2, where it was disclosed that the majority of respondents felt that their particular VREs were developed around their topics and were not specifically driven by the technologies themselves.

Another characteristic listed in the literature study in 2.2.8.3 was that a VRE should render an effective, personalised access point to information, knowledge, collaboration tools, computational resources, and experts. This characteristic was confirmed in the empirical study in 7.3.1.1, Question 5, where it was mentioned that the VRE provides a space that one can access when one needs it. It was also described as a cloud or database, in which people can share articles and information with each other on an aspect/topic that interests all in the group. It was also described as a source that allows interaction and collaboration with people (which could include experts) through a central point as a one-stop solution, where they have all their data and tools.

As mentioned earlier, the literature study identified vertical component layers, namely a management services component, a standards, specifications and protocols component, and a policy component (see 3.5.7.4, 3.5.7.5, and 3.5.7.6). The management services component (see 3.5.7.4) was shown to confer automatic behaviour (which is essential) to the whole VRE across the different layers and components by utilising standards, protocols and specifications in service invocation. None of the respondents in the case studies mentioned this component, and the reason for this could be because this component operates automatically, unseen and unnoticed in the background.

In the discussion on standards, specifications and protocols component in 3.5.7.5, it was mentioned that the various sub-layers within the software components layer are held together by interoperable standards, protocols and specifications, which also help the various software components to communicate with each other and to exchange data with one another. The discussion on protocols indicated that protocols controlled the exchange of data between entities through a set of rules or conventions. Protocols were also shown to comprise data format, signal levels, control information coordination, error handling, and timing. Examples of protocols that were given are the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), Z39.50, and SRU/SRW. Standards and specifications, on the other hand, were shown to be a set of rules, conditions or requirements that prescribe definitions of terms; classification of components; specification of materials, performance or operations; outlining of procedures; or evaluation of quantity and quality in describing, products, services, systems, or practices. In the discussion in 3.5.7.5, various types of standards were named:

- Java standards for programming language technology, classes and standard patterns, e.g. JSR 168 (portlet-1), JSR 286 (portlet-2) and JSR 170 (repository);
- Browser-based web technology standards, e.g. AJAX, CGI, JSP, JavaScript and Portlets;
- Web services standards e.g. SOAP, WSDL, WSRP, UDDI, XML and pub-sub pattern;
- Security standards, e.g. TLS, SSL, Kerberos, GSI, SAML and X.509;
- Metadata standards e.g. MARC and Dublin Core;
- Database management standards, e.g. SQL, JDBC and Hiberbate;
- Data discovery access standards, e.g. Z39.50, OAI-PMH, SRW/SRU, OpenURL and OpenSearch; and
- Workflow standards, e.g. SCUFL and BPEL.

An example of a specification was shown to be API (Application Programming Interface), which can be defined as an “interface (consisting of pieces of programming code) implemented by an application that allows other applications to communicate with it”

(Kashyap, 2010). The VRE-D indicated in the empirical part of the study in 7.3.1.3, Question 42 that he wrote JavaScripts to make some adjustments and customisations to the system. In 7.3.1.1, under Question 8, it was also mentioned that the Alfresco VRE has a Dublin Core Metadata template, designed in line with the Dublin Core Standard. Unfortunately, none of the respondents in either case study had made use of this. The VRE-D further stated in 7.3.1.3, Questions 34 and 35, that he had used open source platforms that were based on Java, PHP, MySQL or PostgreSQL, for the two case studies. In 7.3.1.1, Question 15 (i), the VRE-D revealed that DSpace and Fedora-based systems use open web APIs, which made it possible to get access to the code that would allow another system such as Alfresco to post data into the repository system. He also mentioned in 7.3.1.1, Question 15 (j), that Alfresco has an open API that will enable the plugging-in of a preservation component (system) into the platform. The open API in Alfresco will also make it easier to plug-in several of the pluggable VRE components that were identified.

In 3.5.7.6, the researcher of this study mentioned that every VRE has a number of important policy components, which have to be considered to ensure the successful operation of the VRE. The close relationship between these policy components and the human components layer was also pointed out, as well as their impact on that layer, the functioning of the other layers, and the choice of components used. A list of fourteen potential policy components was also given in Figures 3.12b and 5.2e. The empirical study in 7.3.1.1, Question 19, however, revealed that there had been no formal policies in place in either case study. They did, nevertheless, have mutual agreements on various policy components. The fourteen policy components will be discussed next.

The first policy component mentioned in 3.5.7.6 was to have clear ground rules, and the example given was a decision on who would act as facilitator. In 7.3.1.1, Question 10, it was revealed that both case studies had designated members that acted as facilitators. In 7.2.1.3, it was decided that members of Case Study A would take responsibility for managing their own data.

The second policy component listed in 3.5.7.6 dealt with determining the roles in the VRE. The empirical study in 7.2.1.7 and 7.3.1.1, Question 2, revealed that the roles and rights of the members were clearly defined, and that each member had a clear idea of his/her role within their specific VRE.

Trust relationships was the third policy component mentioned in 3.5.7.6. This component, however, was not mentioned directly in the empirical study, but had to be present to ensure the effective functioning of these VREs.

The fourth policy component listed in 3.5.7.6 was clearly defined objectives. This policy component was elaborated upon in the empirical study in 7.3.1.1, Question 21, and was divided into immediate objectives and objectives beyond the project period. The immediate objectives consisted of the following: use the VRE for storage/back-up of data; use the VRE for retrieval and access to data; use the VRE as a common centralised space/platform; use the VRE for data-sharing; track the progress of student researchers through the VRE; provide access to protocols of experiments, via the VRE; preserve data through the VRE; work collaboratively through the VRE; use the VRE to provide a secure space for one's data; replace paper-based processes with online processes provided through the VRE; and use the VRE as a communication platform (see 7.3.1.1 under Question 21(a)). The objectives beyond the project period consisted of the following**:** provide data processing through the VRE; analyse data through the VRE; create capacity in the VRE to enable it to handle big data sets; add processing and computing power to the VRE; use the VRE for multidisciplinary research; establish a formal structure for RDM at the University; establish a VRE system with an interface similar to Facebook; establish a high performance-computing set-up; and migrate the VRE to another software platform.

The fifth policy component mentioned in 3.5.7.6 was a mutually agreed project plan / collaborative agreement. The empirical study in 7.3.1.1, Question 19, revealed that Case Study A had a mutually agreed project plan/collaborative agreement on how data should be uploaded onto, and archived in the VRE, as well as shared through the VRE, while Case Study B had protocols that guided its members' actions. Protocols that guide members' actions could thus be added to the list of potential policy components.

The handling of intellectual property issues across country borders was the sixth policy component that was mentioned in 3.5.7.6. In South Africa, information (and data) are protected by the Protection of Personal and Information (POPI) Act, which regulates the sharing and storing of personal information across country borders. The empirical study in 7.3.1.1, Question 16 (b), disclosed that there might be some intellectual property issues with the data of members in Case Study A, which could have legal implications. In Case Study B, only one of the members indicated that there were some legal restrictions (intellectual property issues) with her data. The issue of handling intellectual property issues across country borders was not mentioned, however.

The sixth policy component mentioned in the literature study in 3.5.7.6 was protection of rights. None of the respondents in the empirical study mentioned the protection of rights, but they nevertheless mentioned in 7.3.1.1, Question 16 (a), that their data had been restricted for confidentiality and ethical reasons, which alludes to the protection of the rights of their research subjects.

The consideration and handling of ethical issues was the seventh policy component mentioned in 3.5.7.6. The importance of ethical considerations was addressed in 7.3.1.1, Question 16 (a), where it was revealed that the respondents in both case studies had indicated that their data had been restricted for confidentiality and ethical reasons.

The eighth policy component in 3.5.7.6 is the proper matching of skills levels and research interests. The empirical study did not touch on this policy component, but the assignment of roles and rights of members in 7.2.1.7 would have taken into account the skills levels and research interests of members, when the assignments were made.

The decision on type of interface, type of grid service, and/or cloud service, pluggable components, standards and protocols, were the ninth policy component that was listed in 3.5.7.6. The issue of having a type of interface that is more intuitive was mentioned in 7.3.1.1 under Question 21 (c)(vii). The decision on the type of interface platform to be used was taken in 7.2.1.5 (a) and (b). The empirical study revealed that none of the case studies used a grid service or cloud service, but the possibility of considering the use of a cloud service as storage in the future was proposed in 7.3.1.1, Question 21 (c)(viii). The decision on which pluggable components to use was determined by the needs of each VRE case study, and links up with the earlier discussion on the fourth layer of a VRE consisting of RDM components and the bottom layer in the software components layer that comprise various software components that can be plugged into the interface / platform component. The decision on which standards, specifications and protocols components to use would also be dependent on the needs of each VRE case study and links up with the earlier discussion on the vertical layer of standards, specifications and protocols components.

The tenth policy component mentioned in 3.5.7.6 was negotiations/decisions on shared access to publications and conference papers (licensing issues). This policy component was not touched upon in the empirical study, but is an important issue that will need to be considered when providing access to articles, conference papers, etc. within the VRE.

Negotiations/decisions on shared access to research equipment, instruments, and technology was the eleventh policy component given in 3.5.7.6, but although equipment, instruments and technology were mentioned in the empirical study in 7.3.1.1, Question 8, none of the respondents mentioned the aspect of shared access to these. However, the plugging-in of these instruments was suggested as a future possibility, which would suggest shared access by all the members in the VRE.

The twelfth policy component mentioned in 3.5.7.6 was negotiations/decisions on shared opportunities for publishing and presentations. The empirical study in 7.3.1.1 under Question 15 (i) revealed that, at the time of the interviews, none of members of either case study had reached the publishing stage yet, and could therefore not give their views on this. The VRE-D, nevertheless, proposed in 7.3.1.1, Question 12, that decisions on publishing of data should be included in a DMP.

Regular progress monitoring was the thirteenth policy component listed in 3.5.7.6. The empirical study disclosed that monitoring progress was applied in both case studies. This was mentioned by A-R5 in 7.3.1.1, Question 19, as well as by the B-VRE-M in 7.3.1.1, Questions 25 and 30, as part of her tasks as VRE manager. Furthermore, in 7.3.1.1, Question 8, it was revealed that the A-VRE-C used the ‘following’ component to monitor the progress of members of Case Study A.

The aim of a VRE with its different components (human, hardware, software, standards, protocols and specifications, management services and policy) is to support and enhance the research cycle and each of its stages. The following section will touch on this.

8.2.4 How Does A VRE Support A Research Cycle?

In 3.4.1 and in Figure 3.4, the researcher proposed a research cycle for this study, consisting of the researcher's adapted version of Pienaar and Van Deventer's (2009), and Van Deventer et al.'s (2009) research cycle. It contains the following stages, which function iteratively: identification of research area; literature review and indexing; identification of collaborators; proposal writing; identification of funding sources; experimentation and analysis; writing up results; and dissemination / output of findings. In the literature study in Table 3.2, a list of possible VRE components relevant to this study was compiled and matched to the stages of the research cycle. The empirical study also disclosed a number of VRE components that could be matched to the stages of the research cycle, as set out below.

- **Identification Of Research Area**

In this stage, authentication using an authentication service was identified as a component in the literature study (see 3.4.2). The empirical study also identified authentication as a component, and revealed that authentication was done through a username and password (see 7.2.2.5). The A-VRE-C controlled the authentication and access permissions for members of Case Study A (see 7.3.1.1, Question 25), and those of Case Study B was controlled by the B-VRE-M (see 7.3.2).

Another component mentioned in the literature study was personal networks, where the human component communicates and collaborates with colleagues, and maintains awareness of who is currently doing what (see 3.4.2). The empirical study, however, did not confirm personal networks as a component that could be used for identification of a research area.

The literature study further mentioned hypothesis formulation (see 3.4.2), as a component in this stage of the research cycle, but the empirical study did not confirm this.

The next component mentioned in the literature was literature search (to discover what resources are available, research-related information, tracking of research activity and achievement) (see 3.4.2). The literature search component was confirmed by the B-L in the empirical study, when she described her role within the VRE in 7.3.1.1, Question 2, as doing information searches for the student researchers, and uploading the results (articles) onto the VRE platform. Literature searching as component was also mentioned by B-R2 when she indicated in 7.3.1.1, Question 11, that she conducted literature searches and uploaded the articles together with all her search strategies onto the VRE platform. The processes of literature search strategies and selection of appropriate resources, as well as the upload of these, were also mentioned by the B-VRE-M in 7.3.1.1, Questions 11 and 15 (a)(v).

The final component mentioned by the literature study in 3.4.2, which could be matched with the identification of research area, is funders (who provides research related information); nevertheless, none of the respondents in the empirical study mentioned funders as a component, which in the opinion of the researcher is an essential component when starting a research project.

- **Literature Review And Indexing**

The components matched with this stage during the literature study in 3.4.2, comprised a literature search function and referencing. The empirical study confirmed the literature search component through the B-L's response in 7.3.1.1, Question 2, when she described her role within the VRE as doing information searches for the student researchers, and uploading the results (articles) onto the VRE platform. Literature searches as component was also mentioned by B-R2 when she indicated in 7.3.1.1, Question 11, that she conducted literature searches and uploaded the articles together with all her search strategies onto the VRE platform. The processes of literature search strategies and selection of appropriate resources, as well as the upload of these, were also mentioned by the B-VRE-M in 7.3.1.1, Questions 11 and 15 (a)(v). The empirical study further disclosed that the Alfresco VRE did not have a referencing component, with the result that B-R2 recommended in 7.3.1.1, Question 9, that a referencing system (component) such as EndNote or RefWorks should be plugged into the system.

- **Identification Of Collaborators**

In 3.4.2, personal networks (containing the human component and issues of trust, of who will take leadership, and transparency and clarity, communication and collaboration with colleagues, and awareness of who is currently doing what) was shown to be a component that could help in the identification of collaborators. The empirical study in 7.2.1.1 revealed that the members of the Library Services (the researcher of this study, and a member of the executive team of the Library) identified an institute with student researchers in a specific Faculty, which could potentially become a VRE pilot project. A site was created for the group on Moodle on 16 May 2013 and issues such as roles and rights of members were clearly defined on 1 October 2013, as mentioned in 7.2.1.7. The whole idea of personal networks came to the fore in Case Study B, in 7.2.2.1, when the promotor/supervisor (B-VRE-M) of Case Study B contacted the researcher of this study, indicating that she had heard through her personal networks about the VRE used in Case Study A. She expressed her need for a VRE using similar software for her research group, consisting of collaborating student researchers.

- **Proposal Writing**

The literature study in 3.4.2 matched word processing and document management as possible components to the stage of proposal writing. None of the respondents in the empirical study, however, touched on components for proposal writing, although tools for writing up articles, document management, etc. could potentially be used in the writing of proposals as well.

- **Identification Of Funding Sources**

In this stage, the most important component identified in the literature study in 3.4.2 is the identification of funders and funding opportunities. However, none of the respondents in the empirical study mentioned funders as a component, even though this is an essential component when starting a research project.

- **Experimenting And Analysis**

One of the components identified in the literature study in 3.4.2, which could be applied during this stage, is HPC (especially invoking a computation). The empirical study revealed in 7.3.1.1, Question 21 (c)(iv), that the Alfresco VRE did not have the computing power or storage power necessary to invoke a computation. A-R3 and A-VRE-C mentioned that the VRE platform would need, as a future objective, more processing power and more storage space. This was subsequently confirmed by the A-VRE-M, who mentioned that they would need more computing power (see 7.3.1.1 under Question 21 (c)(iv)). The Alfresco VRE also did not have a HPC setup. Actually, the need for the establishment of a high-performance set-up in the VRE in the future was mentioned in 7.3.1.1, Question 21 (c)(vii). HUBzero was mentioned in 7.3.1.1, Question 15 (a)(ii) by the VRE-D as an example of a VRE platform that has HPC built into it.

The literature study further identified the management of intermediate research results, through RDM services. For example, data analysis software could be applied in this stage as a component. Experimentation, simulation and data visualisation were also identified as possible components in this stage. The empirical study in 7.3.1.1, Question 15 (e) and (f) disclosed that processing and analysis of data could not be done through the VRE at the time of the interviews. Experimentation, simulations and data visualisation could also not be done within the Alfresco VRE (see 7.3.1.1, Question 15 (a)(ii) and (iii), and g) and would need to be plugged into the system. Data validation was another component mentioned in the literature study, which could be applied in the experimenting and analysis stage, but none of the respondents mentioned this in the empirical study.

- **Writing Up Results**

Word processing, spreadsheets, presentation software, document management and social media were listed in the literature study in 3.4.2 as components that could be used to assist in this stage, to write up results. The empirical study in 7.3.1.1 divulged that two of the respondents in Case Study B indicated that they had been using MS Word and MS Excel to write up and document their results, after which they synchronised these files, via the Desktop Syncing function, to the VRE. The process of document management was mentioned in 7.3.1.1, Question 8, where it was indicated that members uploaded and downloaded documents to and from the VRE in the Document Library component. With regards to using social media, the student researchers in Case Study A already indicated in 7.2.1.5 during the demonstration of the prototype on 19 July 2013, that they did not need many additional features or trimmings such as social media. It was also mentioned in 7.3.1.1, Question 8, that none of the respondents in either case study had used the Publishing Function of Alfresco, because one could only publish to social media. It was also revealed under Question 8 that the reason for non-use of social media was that both case studies, and especially Case Study A, had not yet developed the social features of the VRE much. It would seem that they had been viewing this as a nice-to-have. Furthermore, none of the respondents mentioned anything about preservation software.

- **Dissemination / Output Of Findings and Closure / Continuation stage**

Two closely related VRE components that could be applied in this stage, according to the literature study in 3.4.2, are publishing (publish outputs, informally through blogs or wikis and formally through conference or journal papers) and archiving (to an online research output repository). The empirical study revealed that none of the respondents had at the time of the interview, reached the stage where they had published articles, papers or data formally through journals or conference proceedings, or archived these in a research repository or data repository (see 7.3.1.1, Questions 8 and 15 (i)). Alfresco also did not have the capacity for publishing other than in social media (see 7.3.1.1, Question 8). These components would need to be plugged into the VRE. Respondents had, at the time of the interviews, also not published anything informally through social media. The reasons for this had been that they had not yet developed the social features of the VRE much, and it would seem that they had been viewing this as a nice-to-have (see 7.3.1.1, Question 8). Another component identified in 3.4.2 was the long-term preservation and management of research results through data curation and management (archive output data and runtime data). The empirical study in 7.3.1.1, Question 12, however revealed that none of the respondents in either case study mentioned the preservation of data, although one respondent did mention that one should ensure that data are not lost, which could allude to data preservation. Yet another component mentioned in 3.4.2, which could be applied in this stage, is peer review. The empirical study showed in 7.2.1.8 (b) that Alfresco had a very good workflow management system that could be used for peer review, but it was only used by the B-VRE-M in her role as supervisor, to review and moderate the student researchers' work. This stage could also be described as the closure stage of the VRE, as mentioned in 3.4.1, but this is not true in all instances. In some instances, the research lifecycle could be continuous, where a research project continues to operate indefinitely. It is foreseen that both the case studies investigated in this study could continue to exist for an indefinite time period, but the software used for the VRE frameworks in each of these studies could potentially be replaced by something else in the future.

8.2.5 What Is RDM?

A discussion on the concept RDM in the literature study in 4.2.3 revealed that there is a variety of concepts related to RDM that are sometimes used synonymously with RDM. These include: data curation, data stewardship, data governance, data archiving, and data management. In 4.2.3.7, it was also shown that RDM could be seen as an overarching concept, and the other concepts - data curation, data stewardship, data governance, and data archiving, as subsets. It was furthermore pointed out that the process of data curation adds value to the data, while someone takes responsibility for the data sets and its tactical function through data stewardship. Data governance was shown to comprise the goals, policies, shared decision-making, planning, strategies, and processes followed, while data management was shown to be more focused on data within an organisational context. In addition, the literature study in 4.2.3.6 revealed that there are various definitions of the concept RDM. The researcher therefore synthesized aspects from these definitions in the following definition:

> RDM is the process of controlling and organising the data generated during a research project, and covers the entire data lifecycle, which includes the planning of the investigation, conducting the investigation, storage and backing up of the data as it is created, preserving the data long-term after the research investigation has concluded, and making the data accessible for future use.

The literature review, as mentioned in 8.2.4, also identified the management of intermediate research results in the experimenting and analysis stage of the research lifecycle, through RDM services, for example data analysis, data visualisation, and data validation, etc.

The empirical study in 7.3.1.1, Question 12, revealed that the majority of respondents in Case Study A focused on storage and accessibility of data in their description of the concept of RDM, which corresponds to the concept of data archiving (discussed in 4.2.3.4), a subset of RDM, and also with the definition of RDM given in 4.2.3.6, which mentions the storage and backing up of the data, as the data is created, and making data accessible for future use. None of the respondents in Case Study A mentioned the research data lifecycle. The responses received from respondents in Case Study B in 7.3.1.1, Question 12, revealed that the majority of respondents described RDM only in terms of storage of data, which also corresponds to the concept of data archiving (discussed in 4.2.3.4), which is only a subset of RDM. The B-VRE-M and the VRE-D, however, described RDM in terms of the RDM lifecycle, which correlates with the definitions given in the discussion on data curation in 4.2.3.1, where it was mentioned that data curation, a subset of RDM, is the active and ongoing activities that data stewards engage in, to add value to research data throughout its entire lifecycle. It also corresponds with the discussion on the research data lifecycle in 4.5.

None of the respondents from either case study mentioned the concept of data preservation in their definitions of what RDM is, although A-R1 did mention that one must make sure that data are not lost, which could be deduced as referring to preservation. The respondents' answers to the question on data preservation in 7.3.1.1 under Question 15 revealed a total misconception of what data preservation is. They had the idea that data uploaded, stored and backed-up on the VRE meant that the data were preserved long-term. This means that some training of these respondents with regards to long-term preservation would be needed.

8.2.6 Why Should A VRE Be An Essential Framework For The Management Of Research Data?

In the literature study in 5.3, the researcher of this study discussed why a VRE could be seen as an essential technological framework or tool for the management of research data.

The definition of Carusi and Reimer (2010: 13), which was presented in 5.3, describe VREs as providing access to data, tools and services through a technological framework. This aspect of provision of access was confirmed by the majority of respondents from both case studies in the empirical study in 7.3.1.1, Question 13. The literature study in 5.3 also revealed that VREs could facilitate collaboration between researchers in the management of data (Brown, 2013; Carusi and Reimer, 2010: 10, 13, 19-20; JISC, 2006; Robertson Library, n.d.). In addition, it was shown in 5.3 that the collaborative nature of VREs afforded opportunities to share data and work together in collecting, manipulating, analysing and interpreting data (JISC, 2006; Carusi and Reimer, 2010: 20). This aspect of collaboration, however, was not mentioned by respondents during the empirical study, and reflected the fact that the social aspect of the VREs in these case studies had not fully developed yet at the time of the interviews (see 7.3.1.1, under Question 8).

The discussion in 5.3 also showed that VREs could be used for the sharing of data with peers, which was shown to be a key element in a VRE (Carusi and Reimer, 2010: 19; Filetti and Gnauck, 2011: 237). This aspect of sharing data among peers was confirmed by responses received from the empirical study in 7.3.1.1, Question 13, where it was mentioned that the VRE afforded members the possibility to share data among members of the VRE. It should nevertheless be emphasized that Masters and Doctoral students normally do not want to share their data before they have been awarded their degrees, and this influenced the restrictions placed and the abnormalities displayed in the two case studies.

The literature study in 5.3 further disclosed that VREs could provide an easy-to-use platform for the short-term storage of their data (Carusi and Reimer, 2010: 18-19; Neuroth, Lohmeyer and Smith, 2011: 225). This was confirmed by the results received in the empirical study in 7.3.1.1, Question 13, where the majority of respondents mentioned that they were using the VRE to save, back-up, store or archive their data.

VREs were also revealed in 5.3 as safe places where researchers can secure their data and keep control of their work (Brown, 2013; Robertson Library, n.d.). This aspect of using the VRE to secure data safely was confirmed by two respondents in Case Study A and one respondent in Case Study B, as well as the VRE-D, in the empirical study in 7.3.1.1, Question 13.

The discussion in 5.3 further disclosed that VREs are ideal technological frameworks for providing excellent access points to repositories, because VREs are typically designed around the researchers' workflow, and are also fully integrated with the research process (cycle) (Carusi and Reimer, 2010: 18; Neuroth, Lohmeyer and Smith, 2011: 223, 230). None of the respondents in 7.3.1.1, Question 13, mentioned repositories. The reason for this became clear in the empirical study in 7.3.1.1, Question 15 (i), where it was divulged that none of the student researchers from either of the case studies had yet published their data in a repository at the time of the interviews. The VRE-D further revealed that at the time of the interviews, it was not possible to publish data through the Alfresco VRE directly onto a repository. He indicated that it would be possible to customise the code in Alfresco, to enable it to publish data from it into a repository. In 7.3.1.3, Question 44, the VRE-D also indicated that one of the future developments for the VRE he foresaw, was the capability of publishing directly from the VRE into a repository.

The discussion in 5.3 furthermore disclosed that the usage of a VRE would be advanced, if it arrived with "a well thought out data management plan and the tools" necessary "to use and create data in documented formats" (Carusi and Reimer, 2010: 18). The idea of compiling DMPs to enhance the usage of VREs, however, was not mentioned by any of the respondents in the empirical study, which is a real shortcoming.

The discussion in 5.3 also revealed that VREs could provide access to data, including to co-researchers that are geographically spread out (Carusi and Reimer, 2010: 18; Neuroth, Lohmeyer and Smith, 2011: 223, 230). In other words, VREs provides access to geographically dispersed researchers to work together on a project and its data

(Carusi and Reimer, 2010: 22). This idea of sharing data with others who are geographically dispersed, were subsequently confirmed by two of the respondents in the empirical study in 7.3.1.1, Question 13.

It was further emphasized in 5.3 that the interdisciplinary nature of VREs made them ideal for the gathering together of data and approaches from different disciplines to create new research findings (Carusi and Reimer, 2010: 23, Fraser 2005). The results from the empirical study in 7.3.1.1, Question 3, mentioned that the two case studies focused on two different disciplinary areas. Case Study A had been using natural science-oriented data, and laboratory/experimental methods, whereas Case Study B had been using human-oriented data and survey instruments as data collection methods. The results from the empirical study in 7.3.1.1, Question 3, further revealed that most of the respondents in Case Study A identified their VRE as a VRE that catered for more than one discipline. One respondent even described it as one discipline cutting across other disciplines or fields. In Case Study B, there were also two respondents that were involved in multi-disciplinary projects. The aspect of using multi-disciplinarity to create new research findings was not mentioned by respondents, however.

VREs were shown in 5.3 to provide a rich environment for the necessary context and provenance that will ensure the trustworthiness of data (Carusi and Reimer, 2010: 42). The Alfresco VRE includes many components that provide context and provenance to data, such as adding metadata, tags, categories, and comments (see 7.3.1.1, Question 8). Results from the empirical study in 7.3.1.1, Question 8, revealed that the student researchers did not add metadata to their data, which showed a lack of knowledge about the value of adding metadata. Similarly, the use of tags, categories and comments revealed a very low usage, which could in future be problematic when trying to understand the context and provenance under which a data file had been created. In addition, VREs were shown in 5.3 to provide the ideal technological frameworks where research data generated through models / simulations, observations, and experiments could be linked with the data collection methodologies and instrumentation (Martinez-Uribe and Macdonald, 2009: 311). The empirical study in 7.3.1.1, Question 15 (ii), divulged that simulations had not been part of the core functions of the Alfresco VRE, with the result that respondents had not been able to run simulations within the VRE. The answers received from respondents revealed that simulations were not of a high priority to them. It was further mentioned in 7.3.1.1, Question 15 (a)(ii), that simulation tools enabled one to do simulations of experiments, and in the process, generate data. The VRE-D revealed that a VRE platform such as HUBzero had HPC built into it, which would enable the possibility to run simulations. In 7.3.1.1, Question 11, the B-VRE-M linked research data collected through observations (a type of collection method) with technology (which could very possibly include instruments). In addition, the empirical study in 7.3.1.1, Question 15 (a)(iii) disclosed that wet lab experiments were not possible within or through a VRE, but that data flowing from these experiments had been uploaded onto the VRE. The discussion in 7.3.1.1 (a)(iii) further emphasized that experiments done via computerised instruments would be possible through a VRE, if these instruments were linked to, or plugged into a VRE. This functionality, however, had not been available in the VRE to respondents in either case study, at the time of the interviews. It was also shown in 5.3 that VREs could further provide researchers with new forms of data and challenges to analysis (Wilson et al., 2007: 290). None of the respondents, however, mentioned that VREs could provide researchers with new forms of data.

The discussion in 5.3 furthermore revealed that a key characteristic of a VRE is that it affords researchers and research teams with more effective ways, as well as the necessary tools, for collecting (capturing), manipulating, managing and securing data collaboratively (Brown, 2013; Robertson Library, n.d.). In 5.4 and 4.5.2, the researcher identified a number of data capture / collection tools that could be added to a VRE technological framework, namely observations, textual or visual analysis, interviews, focus group interviews, surveys, tracking, experiments (using laboratory instruments), sensor instruments, case studies, literature reviews, questionnaires, etc. Observations, experiments, and instruments were discussed earlier in this chapter. In 7.3.1.1, Question 15 (a)(v), the majority of the respondents indicated that they had used the VRE to capture articles that they use as data. The majority of respondents in Case Study B indicated in 7.3.1.1, Question 15 (a)(iv), that they used the VRE to capture survey data. In addition, the VRE-D revealed that Alfresco didn’t have a survey tool built into it, but that he had developed a survey platform (created on LimeSurvey), which he had then plugged into Alfresco (see also 7.2.2.21 (d)).

The discussion in 5.3 further divulged that VREs could be used for analysis and processing of data, as well as for annotating data collaboratively (Carusi and Reimer, 2010: 19; Filetti and Gnauck, 2011: 237). In 8.2.1 above, it was mentioned that the Alfresco platform used for the two case studies could not be used for analysis and processing of data, but that members expressed their need for this component to be added and integrated into the VRE platform in the future. The VRE-D also revealed that it was already available within the HUBzero platform, which he was investigating as a potential future VRE platform (see 7.3.1.1, Question 15 (e) and (f)). The aspect of annotating data was addressed through the ‘Comments’ component mentioned in 7.3.1.1, Question 8. The results showed that members of Case Study B had used the ‘Comments’ function more than members from Case Study A.

In addition, the empirical study in 7.3.1.1 under Question 13 disclosed a number of RDM related aspects that VREs could handle, which had not been found in the literature study. These included adding of metadata to data (based on Dublin Core), tagging of data files, versioning of data files, and monitoring of data.

8.2.7 To What Extent Can The Components Identified Through Question 8.2.3 Be Formalised Into A Conceptual Framework (Model) And Where Would RDM As Component Be Placed?

In Chapter 3, under 3.5.7, the researcher of this study proposed a conceptual framework of a VRE and its components. This conceptual framework of a VRE was further enhanced in Chapter 5 under 5.4, by adding RDM to the model. The VRE conceptual framework (model) and its various layers, as identified in 3.5.7 and 5.4, was confirmed in the discussion under 8.2.3. With regard to the components, however, the empirical study disclosed that some of the components that were identified in the literature study were omitted, and a number of new components could be added. The researcher therefore refined the conceptual model to include the changes that flowed from the empirical study. The revised conceptual VRE framework model can be seen in Figures 8.1 (a-d).

Figure 8.1a: Conceptual Model Showing RDM Components

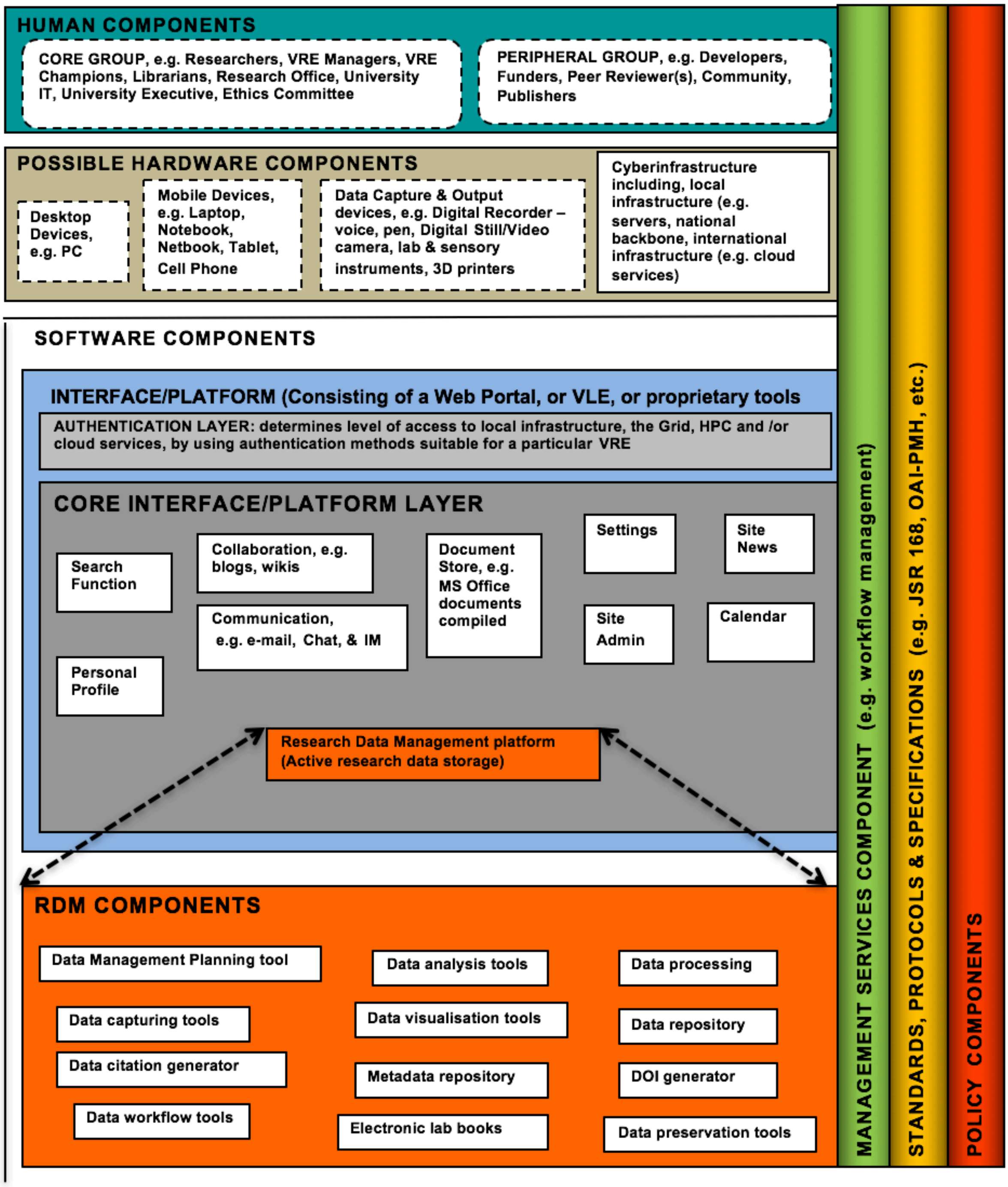
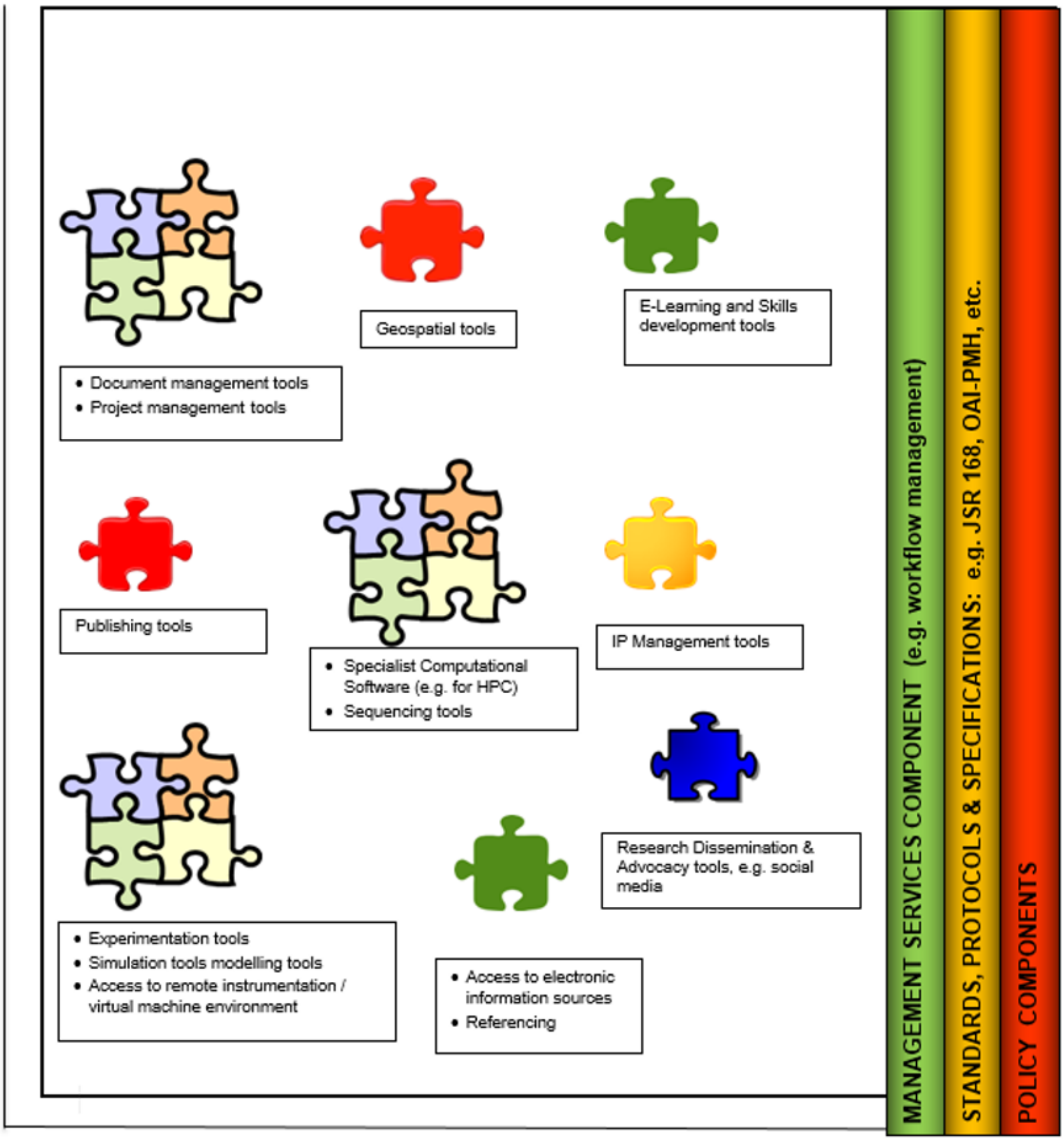

HUMAN COMPONENTS
CORE GROUP, e.g. Researchers, VRE Managers, VRE Champions, Librarians, Research Office, University IT, University Executive, Ethics Committee
PERIPHERAL GROUP, e.g. Developers, Funders, Peer Reviewer(s), Community, Publishers
POSSIBLE HARDWARE COMPONENTS
Desktop Devices, e.g. PC
Mobile Devices, e.g. Laptop, Notebook, Netbook, Tablet, Cell Phone
Data Capture & Output devices, e.g. Digital Recorder – voice, pen, Digital Still/Video camera, lab & sensory instruments, 3D printers
Cyberinfrastructure including, local infrastructure (e.g. servers, national backbone, international infrastructure (e.g. cloud services)
SOFTWARE COMPONENTS
INTERFACE/PLATFORM (Consisting of a Web Portal, or VLE, or proprietary tools
AUTHENTICATION LAYER: determines level of access to local infrastructure, the Grid, HPC and /or cloud services, by using authentication methods suitable for a particular VRE
CORE INTERFACE/PLATFORM LAYER
Search Function
Collaboration, e.g. blogs, wikis
Communication, e.g. e-mail, Chat, & IM
Document Store, e.g. MS Office documents compiled
Settings
Site News
Site Admin
Calendar
Personal Profile
Research Data Management platform (Active research data storage)
RDM COMPONENTS
Data Management Planning tool
Data analysis tools
Data processing
Data capturing tools
Data visualisation tools
Data repository
Data citation generator
Metadata repository
DOI generator
Data workflow tools
Electronic lab books
Data preservation tools
MANAGEMENT SERVICES COMPONENT (e.g. workflow management)
STANDARDS, PROTOCOLS & SPECIFICATIONS (e.g. JSR 168, OAI-PMH, etc.)
POLICY COMPONENTS

Figure 8.1b: Other Pluggable VRE Components



Geospatial tools
E-Learning and Skills development tools
Document management tools
Project management tools
Publishing tools
Specialist Computational Software (e.g. for HPC)
Sequencing tools
IP Management tools
Research Dissemination & Advocacy tools, e.g. social media
Experimentation tools
Simulation tools modelling tools
Access to remote instrumentation / virtual machine environment
Access to electronic information sources
Referencing
MANAGEMENT SERVICES COMPONENT (e.g. workflow management)
STANDARDS, PROTOCOLS & SPECIFICATIONS: e.g. JSR 168, OAI-PMH, etc.
POLICY COMPONENTS

Figure 8.1c: Model Applied To Research Lifecycle And Research Data Lifecycle

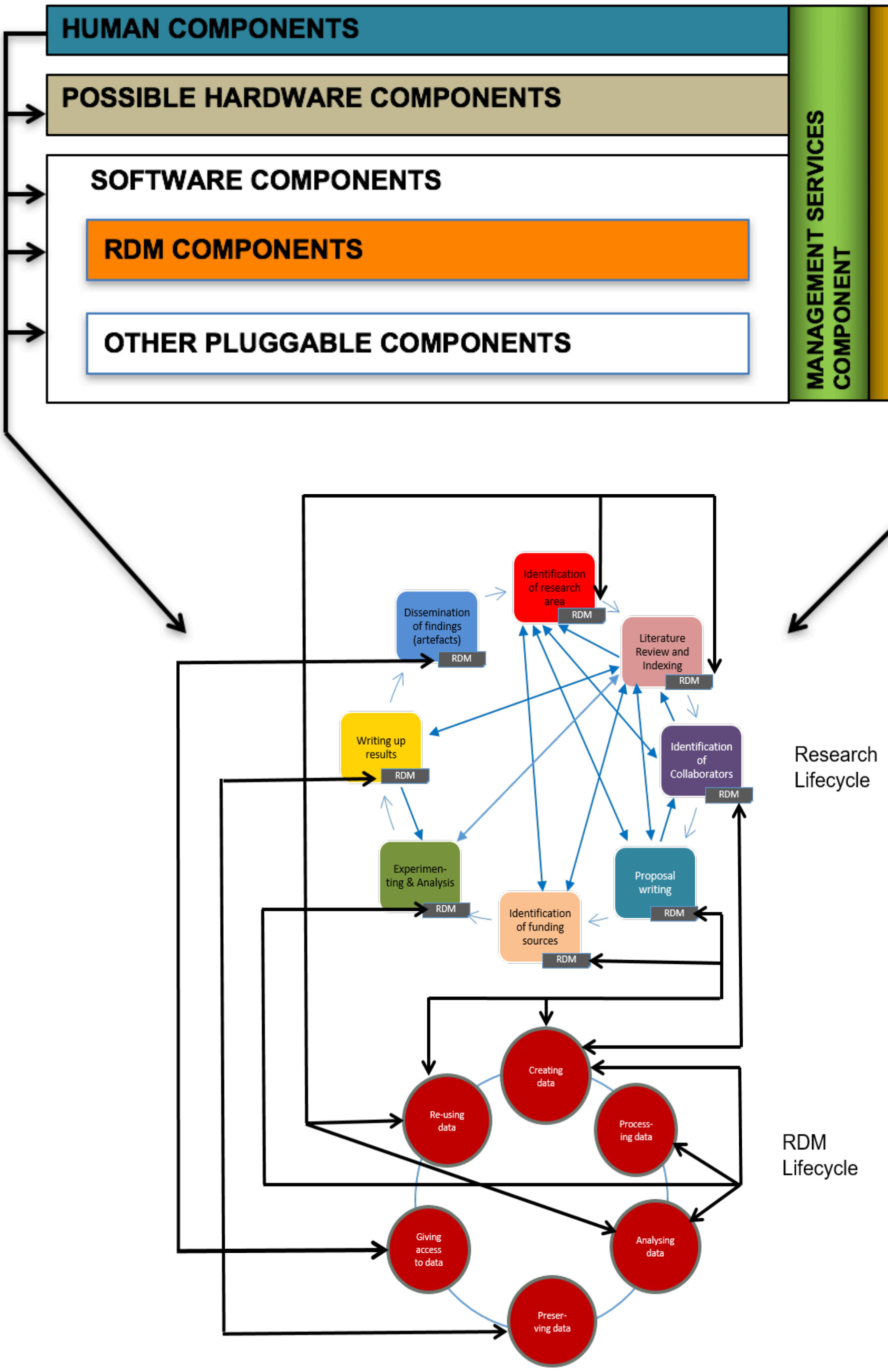

HUMAN COMPONENTS
POSSIBLE HARDWARE COMPONENTS
SOFTWARE COMPONENTS
RDM COMPONENTS
OTHER PLUGGABLE COMPONENTS
MANAGEMENT SERVICES COMPONENT
STANDARDS, PROTOCOLS & SPECIFICATIONS
POLICY COMPONENTS
Identification of research area
RDM
Dissemination of findings (artefacts)
RDM
Literature Review and Indexing
RDM
Writing up results
RDM
Identification of Collaborators
RDM
Experimen-ting & Analysis
RDM
Proposal writing
RDM
Identification of funding sources
RDM
Research Lifecycle
Creating data
Re-using data
Process-ing data
Giving access to data
Analysing data
Preser-ving data
RDM Lifecycle

Figure 8.1d: Research Lifecycle And RDM Lifecycle (Enlarged View)

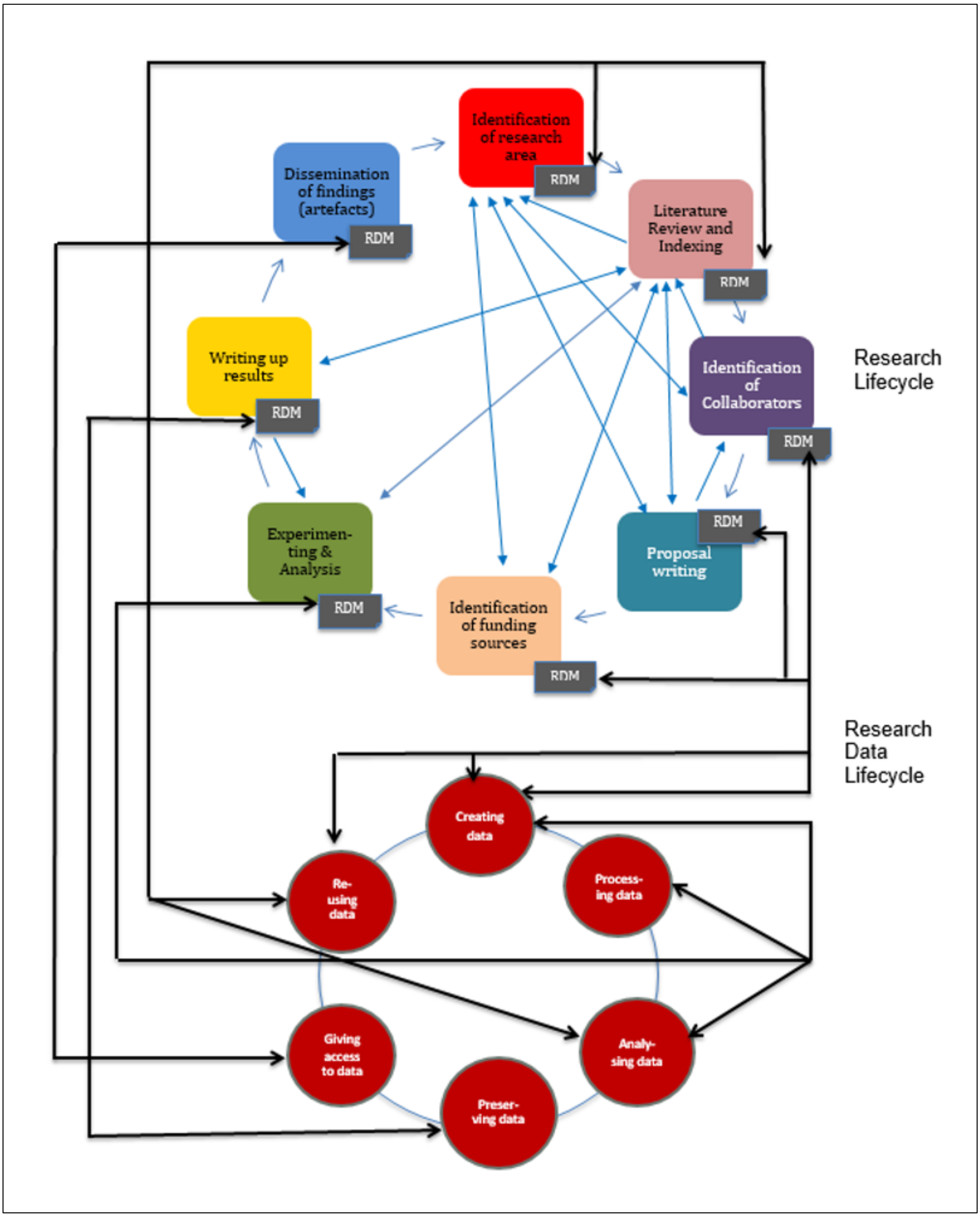


Figure 8.1e: Policy Components (Expanded)

- Clear ground rules, e.g. Determine who act as facilitator; Determine the roles in the VRE
- Trust relationships
- Clearly defined objectives
- Mutually agreed project plan/collaborative agreement
- Encouragement of shared interest and enthusiasm
- Intellectual Property (IP) issues across country borders should be dealt with beforehand
- Protection of rights
- Ethical issues must be considered and taken care of
- Proper matching of skills levels and research interests
- Decision on type of interface, type of grid service, and/ or cloud service, pluggable components, standards and protocols
- Negotiations / Decisions on shared access to publications, conference papers (licensing issues)
- Negotiations / Decisions on shared access to research equipment, instruments, and technology
- Negotiations / Decisions on shared opportunities for publishing and presentations
- Regular progress monitoring

The core group in the human components layer was confirmed, but the Research Office, University IT Services, and the University Executive were not mentioned in the empirical study. These three entities would be important components to ensure the successful functioning of a VRE, and were therefore kept in the model. The Research Office would be involved in policy as well as with assistance to researchers with regard to funders' requirements and DMPs. The University IT Services would ensure and maintain the necessary IT infrastructure (local network infrastructure such as servers and network connections; storage and computing power; HPC; connectivity to the national backbone; and connectivity to the international infrastructure). The University Executive would take the responsibility of seeing to it that the RDM policy of the University is adhered to.

In 8.2.3, the discussion revealed that the empirical study had added the Faculty Ethics Committee as a core component in the human components layer. The peripheral group in the human components layer was also validated in the empirical study. Funders, peer reviewers and the community were not mentioned in the empirical study, but were nevertheless kept in the model. Funders are an important human component and would typically require DMPs, require that research data are well managed, and that data are published in an open source data repository. Peer reviewers are also an important human component. As mentioned in 3.5.7.1, peer reviewing forms an essential part of the research process. They ensure that the data that are uploaded onto a VRE are of a high quality. The data generated through a VRE could be of value to the community and members of the community could be provided access to certain parts of a VRE.

Only three of the categories of the hardware components layer, as mentioned in 8.2.3, were confirmed by respondents in the empirical study, namely desktop services, mobile devices and cyberinfrastructure. Although the fourth category, data capture and output devices, was omitted by respondents, it would still be an important category to include, and the usage of these would depend on the type of project a VRE focused on. The usage of the examples given in each of the categories would also depend on the needs of each project. The empirical study further added laboratory and other instruments as a possible component; these have been included in the data capture and output devices category.

The software components layer as identified and described in 3.5.7 and 5.4 was confirmed by the empirical study, as discussed in 8.2.3. The discussion in 8.2.3 also confirmed the authentication layer as part of the software layer, to determine the level of access a human component can have to the software layer. An additional method of authentication which was not mentioned in the literature study, was proposed by the VRE-D when he mentioned the possibility of replacing Alfresco with HUBzero and adding this to the University's authentication directory, namely active directory, so that any researchers can sign into the VRE, using his / her University credentials.

The discussion in 8.2.3 further mentioned that the software layer includes a core interface layer consisting of fixed components that are part of the standard configuration of the specific tool used. The discussion further indicated that the components used in this core interface could vary, and then mentioned a number of fixed components. Most of these components were confirmed by the empirical study, but instant messaging and chat as communication tools were not mentioned, however, which could be because the social aspect of the case studies not having been developed much at the time of the interviews, as mentioned earlier. These components have therefore been kept as possible components that would be of value in VREs that have developed their social aspects. The respondents also did not confirm the 'settings function' and 'site admin function', which would typically be something that would appear in the interface that the VRE-D would see, but not in all of the respondents' VRE interfaces.

The literature study identified the core interface layer in 5.4 as the place where a RDM component could be situated. The empirical study, however, revealed that respondents used the document management component as a research data store. The reason for this could be related to the tool that was used as a VRE platform. The tool used in the case studies, Alfresco, a document management system (see 7.3.1.3 Question 34), did not specifically contain a RDM store, but did contain a document management store, which was used to store the case studies' data. The literature study in 5.4 mentioned that more components could be added, to enhance the RDM functionality. This led to the identification of a fourth VRE layer, containing RDM components. The majority of the RDM components mentioned in 5.4 were confirmed by the empirical study, as mentioned in 8.2.3. The respondents, however, failed to mention the possibility to add a metadata store, which is in line with the finding in 7.3.1.1, Question 8, that the majority of the members in these case studies did not understand the necessity or value of adding metadata to their documents and data, and therefore would not have mentioned the possibility of adding a metadata store to the VRE. The metadata store component was kept as a component in the model, because of the value it would add to an RDM platform / store. Data management planning tools were also omitted by the respondents as possible RDM tools, and were identified in 7.3.1.1 as an important shortcoming in the case studies. The importance of data management planning tools was emphasized in 7.3.1.1, Question 12, where it was stressed that DMPs indicate what the types of data are that will be generated, what is going to happen to the data, where it is going to be stored, how it is going to be stored, and for how long it will be stored, taking into account the University's or funder's guidelines. A data management planning tool was therefore kept as a component in the RDM components layer of the model.

Most of the pluggable software components identified in the bottom layer in the software components layer, were confirmed through the empirical study, as discussed in 8.2.3. A number of these tools, however, were not mentioned in the empirical study. These are e-learning tools, skills development tools, modelling tools, intellectual property management tools, access to electronic information sources, and a DOI generator. The reason for not mentioning e-learning tools and skills development tools could be because these researchers already underwent training by the VRE-D, and in Case Study A, also by the A-VRE-C; therefore, they possibly felt that they had no need for further e-learning tools and skills development tools. The empirical study, however, revealed in 7.3.1.1, Question 7, that those members of the case studies that had not attended the scheduled training sessions, experienced the VRE differently and struggled with the system. The VRE also kept on developing in line with the needs of the group, with the result that the inclusion of e-learning and skills development tools could be valuable to all members of these VREs. Rapid and continuous developments in the technological environment would, furthermore, necessitate further skills development, which could be addressed by e-learning tools and skills development tools. The researcher therefore kept these components in the model as pluggable VRE tools.

Modelling tools were shown to be used synonymously with simulation tools by Martinez-Uribe and Macdonald (2009: 311). The empirical study in 7.3.1.1, Question 21 (c)(ii), disclosed that there was a need for simulation tools to be plugged into the VRE system, which meant that modelling tools could, by default, be kept as a pluggable VRE component.

The respondents accessed electronic information sources as part of the research process, but none of them mentioned the possibility of having access to these sources within the VRE, and this could be, as mentioned in 8.2.3, related to the fact that they had never considered the possibility. The possibility of having this as a pluggable VRE component was kept, as this would make it possible to support as much of the research lifecycle within one technological framework as possible.

The non-mentioning of intellectual property management tools could be related to the fact that the members of the two case studies had not reached the publishing stage of their research yet, where intellectual property issues would play an important role (see 7.3.1.1 under Question 8). The empirical study revealed further that there might be some intellectual property issues with the data of members in Case Study A, which could have legal implications, while one of the members in Case Study B indicated that there were some legal restrictions (intellectual property issues) with her data. This meant that intellectual property may again arise when these student researchers reached the publishing stage of their research. The intellectual property management tools as a component was thus kept in the model, as a pluggable VRE component.

The non-mentioning of a DOI generator as a possible RDM component, as mentioned in 8.2.3, could be because none of the respondents, at the time of the interviews, had published their data yet. This component would become essential once their data are published, and the component has therefore been retained in the model, as an RDM component.

The discussion in 8.2.3 further disclosed that the literature study had identified three vertical component layers in a VRE, namely a management services component, a standards, specifications and protocols component, and a policy component, which were confirmed to be essential for the successful running of the VRE and the other component layers. It was also mentioned in 8.2.3 that the management services component, although not mentioned in the empirical study, was seen as an essential component layer to ensure automatic action to the whole VRE between layers and components.

In the discussion on standards, specifications and protocols in 8.2.3, it was shown that most of the types of standards listed in the literature study under 3.5.7.5 were confirmed by the empirical study in 7.3.1.3, Questions 34, 35, and 42, except security standards and data discovery access standards. Although these two were not confirmed in the empirical study, the researcher of this study still deemed these standards as important enough to be included in the conceptual framework model, to ensure security of the data and to make data discoverable. None of the examples of protocols, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), Z39.50, and SRU/SRW were mentioned by the respondents; nonetheless, these were seen as essential for inclusion in the conceptual framework model of a VRE, in that it ensures the successful functioning of a VRE. Protocols as mentioned in 8.2.3 specifies data format and handles signal levels, control information coordination, error handling, and timing, and were included in the model as an essential components layer.

The next component layer that was discussed in 8.2.3 was the policy components layer. It was pointed out that this layer has a close relationship with the human components layer and also has an impact on all the other layers within the VRE. The majority of these policy components were confirmed by the empirical study, as discussed in 8.2.3, and have been kept in the model. There were, nevertheless, a number of policy components that were not confirmed by the empirical study. These included: trust relationships, the handling of intellectual property issues across country borders, the protection of rights and indigenous knowledge rights, negotiations/decisions on shared access to publications and conference papers (licensing issues), and negotiations/decisions on shared opportunities for publishing and presentations. In the discussion under 8.2.3, it was mentioned that although trust relationships were not mentioned in the empirical study, it had to be present to ensure the effective functioning of the VRE internally, and was therefore kept as a policy component in the model. The issue of handling intellectual property was confirmed by respondents, but not the issue of handling intellectual property issues across country borders. This issue is something that might become essential as the VREs develop further, and have thus been kept as a policy component in the model. The discussion in 8.2.3 further revealed that the protection of rights of research subjects was alluded to in the empirical study in 7.3.1.1, Question 16 (a). The protection of rights of research subjects have therefore been kept as a policy component in the model. The issue of protection of indigenous knowledge rights were not mentioned, however, and might be something that would only be subject/discipline specific. It was therefore removed from the list of policy components in the VRE model.

In 8.2.3, it was also pointed out that the policy component about negotiations / decisions on shared access to publications and conference papers (licensing issues) was not touched on in the empirical study. The researcher of this study was of the opinion though, that this component is essential in a VRE and should be included in the VRE model, as it is something that will need to be addressed when providing access to articles, conference papers, etc. to other members within the VREs.

The discussion in 8.2.3 disclosed that the component on negotiations / decisions on shared opportunities for publishing and presentations, had not been mentioned by any of the members in either case study, because they had not reached the publishing stage of their research yet, at the time of the interviews. It was also mentioned that the VRE-D had proposed that decisions on publishing should be included in a DMP. The researcher is of the opinion that this policy component would become important when researchers reach the publishing stage of their research, and therefore should be kept in the VRE model.

8.2.8 To What Extent Can This Model Be Generalised For Use In Other Environments?

During the discussion in 3.2.1 of the UK's VRE programme, it was mentioned that the hope with the 1st phase of the UK programme was to bring all the facets in the different UK VRE projects together into one VRE solution in a similar manner as VLEs, using shelf-ready tools such as SharePoint, Blackboard, Sakai, Moodle or uPortal. The results, however, showed that each of the projects had very distinct and different needs with regards to infrastructure and resources, which made it difficult to bring them together into one standardised solution (Interview with F. van Till and M. Dovey, JISC on 1 June 2010 at the HEFC Building, London). An interview with Van Till from JISC in 2010 suggested that it might perhaps be possible to create or use a centralised framework or standardised platform (which can be transferrable), onto which people can build their own tools (Interview with Van Till and Dovey, JISC on 1 June 2010 at the HEFC Building, London, UK).

The VRE conceptual model presented in 3.5.7, adapted in 5.4, and discussed in 8.2.7, showed that the model could potentially be used in other environments, as can be seen in the two case studies, which were from two different disciplinary areas (see 8.2.2). Case Study A focused on natural science-oriented data and laboratory/experimental methods, and Case Study B on human-oriented data, using survey instruments as data collection method. Due the limitations of the study (the study only focused on two case studies), the conceptual model could therefore not be generalised for use in other environments. Nevertheless, guidelines (see 8.6) can be developed for such a conceptual VRE Model, which could then be applied in multiple disciplinary areas/environments.

8.2.9 How Was The Central Research Question Answered?

The central research question was answered within the framework of e-Research, and within the context of the social sciences approach to e-Research.

The researcher defined a VRE as a common, flexible, technological and collaborative framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts etc.) interoperating with each other, can be plugged, to enable collaboration and to support and enhance large and small scale processes of research, which are performed by researchers in multidisciplinary contexts and across organisational and geographical boundaries.

The study showed that VREs are multi-disciplinary in nature, in other words it could be used in all disciplines and all types of research. VREs, in addition, provide interfaces to cyberinfrastructure (the digital side of research infrastructure). The use of VREs was further shown to give rise to organisational transformation, for example VREs made it possible for members to access data from anywhere and to share information, files and data with one another (in other words collaborate) from any geographical location in the world. The study further revealed that VREs are flexible frameworks that could be adjusted to fit user needs. VREs, moreover, provide researchers with more effective means of collaboratively collecting, manipulating, analysing and interpreting data (see 8.2.6). The results of the study, in addition, confirmed that VREs consist of a number of components (including RDM components) that have been joined together to support the various stages of the research data lifecycle and the research lifecycle (research process) (see 5.4, 8.2.1, and Figure 8.1c). Further results verified that VREs could be used for analysis and processing of data, annotation of data collaboratively, and sharing of data with peers. This sharing of data was shown to be a key element in a VRE (see 6.2.6).

The discussion under 6.2.6 further disclosed that VREs provided easy-to-use technological frameworks for the saving, back-up and storage of data in secure places. VREs were furthermore shown to provide access points to data repositories. The study also revealed that the uptake of VREs by researchers could be enhanced if it included data management planning tools. The interdisciplinary nature of VREs furthermore were shown to make them ideal for the gathering together of data and approaches from different disciplines to create new research findings. The study also disclosed that VREs provided the necessary environments to ensure the context and provenance of data. VREs, in addition, were shown to provide the ideal technological frameworks where research data generated through models/simulations, observations, and experiments could be linked with data collection methodologies and instrumentation.

The discussion on the current state of VRE research in the world under 8.2.2, confirmed that VREs in these different countries had similar key elements (components) aimed at facilitating shared use of digital infrastructure by researchers through provision of shared environments. All the countries investigated, in addition, used participative/collaborative design processes, which were user driven. The answers to this sub-question further revealed that the making available of data / data sharing through VREs was an important trend. The discussion under this sub-question, in addition, revealed that VREs were structured in such a manner that they supported the research lifecycle. Another issue that was raised is the provision of funding for infrastructure (in this case for a VRE), which would enable the management of research data. The answers also showed that the sharing of all data (total access to everything) was not always possible, but that some data may have certain usage or access rights, and that librarians could play a role in advising on these.

The study showed that VREs could follow a multi-tier (n-tier) architecture, as shown by the design of a VRE conceptual framework model as discussed in 8.2.7. Such a conceptual model was also shown to consist of generic components, which could be grouped in various layers that interact with each other (see 8.2.3). The discussion under 8.2.7 revealed that each of the human components in the human components layer plays an important role in ensuring the effective management of research data through a VRE. The hardware components layer determines the necessary devices that would meet the projects' needs. The authentication layer determines the level of access a human component can have to the software components layer, and ensures that the system and its underlying data is secure. The results from the study further reveal that a RDM platform (consisting of a data store that would facilitate active data storage), could be placed within the core interface layer, which are situated within the software components layer. The RDM functionality could then be enhanced, by adding a number of components, which were confirmed by the empirical study. These RDM components were listed as data capturing tools, data repository tools, data citation generator, data workflow tools, data processing, data preservation tools, data analysis tools/software, electronic laboratory tools, and data visualisation tools. A data management planning tool, a metadata repository/store and DOI generator were not mentioned in the literature study, but were nevertheless added as essential components (see 8.2.3 and Figure 8.1a). The access to other pluggable tools (see Figure 8.1b), would further enhance the effective management of research data through a VRE, in the context of the research lifecycle.

This discussion of the main elements from the answers to the sub-questions revealed that RDM has a central place within a VRE, and that a VRE provides the necessary framework/environment for the effective management of research data, taking into account all the stages of the data lifecycle, within the context of the entire research lifecycle.

8.3 REFLECTION

The literature and case studies showed that there is huge potential for the successful application of VREs in the research lifecycle. The results also revealed that VREs are ideal technological and collaborative frameworks, that can be used for the management of research data throughout the research data lifecycle.

The non-use of a number of the features (components) could be related to the nature of the software (Alfresco) that was used. For example, the empirical study revealed that some of the respondents felt that the Alfresco platform did not have an intuitive, user-friendly interface, as discussed in 7.3.1.1, Question 21 (c)(vii). Some of these components were, nonetheless, plugged into the VRE by the VRE-D, as the needs for these arose. The future migration from the Alfresco platform to HUBzero, as mentioned by the VRE-D, might eventually solve the problem of non-use.

The study further revealed that there is a need for the establishment of a formal structure for RDM at the University of Pretoria. Furthermore, the literature, as well as the results received from the case studies, revealed the important role that a librarian could play in a VRE, as part of the research group (e.g. as embedded librarian). However, it was also revealed that librarians would need to be enlightened on their potential role within VREs, and would need a considerable amount of training and upskilling, to enable them to be able to play a significant role within VREs. This study disclosed that librarians could, for example, determine the user requirements for the VRE because of their liaison role, which could then be shared with the VRE designer. In some cases, they could even create tools and interfaces that will allow for the searching and usage of the information resources. Librarians could also do information searches and uploads of literature (data collection/capture), add and assist researchers with metadata, conduct e-Research literacy training sessions to train researchers in the use of a VRE and its components, advise researchers on current forms of electronic preservation methods and formats, assist with the curation and preservation of digital assets, assist researchers in organising their file structures and file names according to file naming conventions, train researchers on data citation methods and principles, train researchers on data management planning tools, and advise researchers on publishing, copyright, open access and licensing issues.

8.4 CONTRIBUTION OF THIS STUDY TO THE SUBJECT FIELD

This study contributed in a number of ways to the subject field:

(i) The VRE conceptual model with its various component layers and generic components were compiled, as shown in 3.5.7, adapted in 5.4, and validated through the empirical study, with the final version in 8.1a-d, which was discussed under 8.2.7.

(ii) The important role that VREs play in providing a technological and collaborative framework for the successful management of research data was also revealed during this study, as pointed out in 8.2.6. With regards to the management of data, the literature and empirical study identified a number of ways in which VREs could contribute:

- Provision of access to data, tools and services through a technological framework;
- Facilitating collaboration between researchers in the management of data;
- Sharing of data with peers;
- Offering an easy-to-use technological framework for short-term storage of data;
- Provision of safe / secure places for researchers’ data;
- Provision of access points to data repositories;
- Provision of access to geographically dispersed researchers that would enable working together on a project and its data;

- Allowing for the gathering together of data and approaches from different disciplines to create new research findings;
- Provision of a rich environment (with metadata, tags, categories and comments) for the necessary context and provenance, so as to ensure the trustworthiness of data;
- Provision of a technological and collaborative framework where research data generated through models / simulations, observations, and experiments could be linked with the data collection methodologies and instrumentation;
- Providing researchers with new forms of analysis and challenges to analysis;
- Provision of the necessary tools for collecting (capturing) data;
- Allowing for the annotation of data through the 'comments' component;
- Provision of analysis and processing tools through a VRE;
- Provision of a data versioning function;
- Allowing for the tracking of the workflow (monitoring) of data; and
- The ability to add metadata.

(iii) The important role of DMPs within VREs was pointed out in the literature study, and also by the VRE-D in 7.3.1.1, Question 12; however, DMP tools were not confirmed by the empirical study as possible RDM tools that are part of the normal research lifecycle. This revealed an important shortcoming in these case studies, and is something that will need attention in other VREs.

(iv) A number of facilities, equipment and hardware were identified through the literature review in 3.5.7 and 5.4. The empirical study further revealed a number of potential facilities, equipment and hardware, which were not found in the literature reviews. These included servers, server rooms, NAS (Network Attached Storage) devices, replication of NAS devices to other devices at different geographical locations, network points, desktop computers that had good storage and computing power, as well as HPC set-up. Other related issues that came up are the solving of firewall issues, and the provision of virtual machine environments, multiple levels of security, managed access, and a data access policy, etc.

(v) A potential new role for librarians were identified, as was discussed in 8.3.

(vi) Important role players in both VREs and RDM were identified, as discussed in 3.5.7, 5.4 and 8.2.3.

8.5 LIMITATIONS OF THIS STUDY

This study was limited because:

- Only two case studies were used to verify results gained from the literature study, and this inhibited the possibility to generalise the proposed conceptual framework model that was presented in 3.5.7, and adapted in 5.4, for use in other studies;
- The research was only conducted within an academic context (where individual researchers / students are conducting research to gain a degree and where face-to-face contact is relatively easy). The outcome of this study might have been different had the researchers been based in research organisations where several researchers from a variety of different parent organisations, would all be working on the same project, and from remote locations;
- None of the student researcher members of either of the case studies had reached the publishing stage of the projects, which contributed to the non-use of some of the components. Had they reached the publishing stage, it might have impacted the outcomes of the study;
- The study did not explore other developing country initiatives; and
- The data from these two case studies were not accessible to users from outside the VRE, because of ethical and legal reasons. This inaccessibility is against the principle of open access.

8.6 GUIDELINES AND RECOMMENDATIONS

A number of guidelines could be developed to set up a VRE model. These are:

- Identify a potential research project that can form the basis for a VRE (see 7.2.1.1 and 7.2.2.1).
- Explore the aim, as well as the current research environment of the identified VRE Group / Project (see 7.2.1.2 and 7.2.2.3).
- Engage with potential role players (human components):
- Arrange a meeting with decision makers, e.g. Executive Management, manager/supervisor of project, Research Office, and Ethics Committee (human components) (see 7.2.1.1);
- Arrange a meeting between potential users of the VRE, e.g. researchers, supervisor/manager of the research project, and potential designer(s) /

developer(s) that can design, set-up, and/or customise the VRE according to the group's needs (see 7.2.1.2, 7.2.1.3 and 7.2.1.4);

- Do an analysis of the needs of the members of the VRE group, and in the analysis focus on each of the stages of the research lifecycle (see 7.2.1.4, 7.2.1.5, 7.2.2.4, and 8.2.4);
- Identify a potential software platform that would meet the group's needs. This could be done by customising a LMS, e.g. Moodle or Sakai, or a document management system, e.g. Alfresco, or by acquiring/using a shelf-ready VRE product such as HUBzero or Open Science Framework (OSF). In the evaluation of potential software platforms, investigate the authentication layer and its potential to integrate with the institution's authentication directory (see 7.2.1.2 and 7.2.1.8).
- Create a prototype of the VRE on the chosen software platform, demonstrate this to members of the group, and do some adaptations if needed (see 7.2.1.4);
- Identify which hardware components are already used by members of the VRE group, and which hardware components will be needed for the successful functioning of the VRE, for example desktop devices, mobile devices, data capture and output devices, as well as cyberinfrastructure. In addition, determine if the system has a synchronisation (syncing) ability (see 7.2.1.3 and 7.2.1.5);
- Clearly define the roles and rights of each of the users (human components) of the VRE: e.g. VRE Manager, VRE Champion, researchers, VRE designer, and librarian(s). For example, give the supervisor/promotor and the VRE Champion, and perhaps also the librarian, full VRE Manager rights, and give the participating researchers only rights to access and edit their own spaces and to read and access shared spaces (see 7.2.1.7);
- Register each of the VRE members onto the system, so that they can be authenticated when logging into the system, and integrate this with the institution's (e.g. University) authentication directory, for example the active directory (see 7.2.1.5 (b) and 7.2.2.5);
- Identify which components are available in the core interface of the VRE software platform. If the core interface does not have a RDM storage platform, add a data store or customise the document management store, if available, in the core interface. To expand the functionality of the RDM platform, add those RDM components that will address the needs of the VRE group and support the

research lifecycle stages. Plug other VRE components into the core interface of the software components layer, as needed by the VRE group and which would support the research lifecycle (see 7.2.1.8 and 7.2.1.14);

- An essential RDM component that should be added, is a potential repository tool (platform) for data publishing. This tool can either be an open source tool or a proprietary product. The identification of the most effective tool at the best cost will have to be negotiated with all the stakeholders, including the executive of the institution, the research office, the library, as well as the IT Services Department, because it will have budgetary implications (see 7.3.1.1, 15 (i), and 7.3.1.2, Question 29);
- Implement the software platform (see 7.2.1.8 (c) and 7.2.2.4);
- Conduct a hands-on training session with the members of the VRE (7.2.1.9 and 7.2.2.5);
- Do iterative formative evaluations to see if the VRE complies with all user requirements, and adapt the VRE accordingly;
- Set up a formal policy for the VRE containing the policy components that would be applicable to the specific VRE (see 7.3.1.1, Question 19, and discussion in 8.2.3);
- Be aware that there is a management services component that confers automatic behaviour to the workflow in the VRE (see discussion in 8.2.3); and
- The types of standards, protocols and specifications for the VRE will be determined by the identified software platform used. It will also be determined by the customisations that will be needed to add the RDM and plug-in other VRE components, in order to meet the needs of the human components, and address all the research lifecycle stages (see discussion in 8.2.3).

The researcher of this study would recommend that the designer(s) of a VRE keep record of every interaction (meetings, training sessions, e-mails, and notes) with members of the VRE, as well as a record of changes/adaptations to the VRE system. This would be very valuable if further research on the VRE is done, or if a report needs to be compiled for the University Executive, funders, etc.

Flowing from the discussion in 8.2.5, the researcher would recommend that the development of long-term preservation skills is required.

It is further recommended that several training and awareness opportunities on VREs and RDM are created.

It is also recommended that further in-depth research be conducted on the role of the librarian in VREs and RDM, and that the curriculum for the training of librarians should be revised to prepare librarians for this changing role. This curriculum could, for example, include:

- Determination of user requirements with regard to VRE design;
- Creation of tools and interfaces that will allow for searching and using of interfaces within a VRE;
- Capturing / collection of data;
- Expertise in metadata and metadata schemas;
- The ability to conduct e-Research literacy training sessions and provision of advice in the use of a VRE and its components;
- Data analysis tools;
- Data visualisation tools;
- Data cleansing tools and techniques;
- Electronic preservation methods and formats;
- File naming conventions;
- Data citation methods and principles;
- Data management planning tools;
- Data publishing;
- Copyright and licensing of data; and
- Open data.

In addition, it is recommended that a VRE champion be identified from the start to ensure successful utilization of a VRE. The VRE champion can encourage members to upload data, mentor members, give advice, and conduct training for individual members of a VRE, as needed. It is also recommended that any VRE that is implemented at the University, be integrated with the researchers' research workflow and day-to-day activities.

8.7 SUGGESTIONS FOR FURTHER RESEARCH

The study revealed a number of areas that could provide opportunities for further research:

- The development of a new curriculum for the training of librarians in VREs and RDM could be investigated (see 8.6 for suggested topics that this investigation could include).
- The role of the librarian in the management of big data could be investigated, and this investigation could look at the skills that librarians would require to assist in the management of big data, for example data curation, indexing and abstracting skills, understanding of metadata and taxonomy, data mining, data visualisation, digital preservation, collaboration, teaching, and facilitation.
- The role of VREs in the management of big data could be researched, and could, for example, look at:
  - How big data could be captured through a VRE;
  - The types of metadata and the challenges of describing big data;
  - The preservation of big data;
  - The complications in citing big data; and
  - The publishing, copyright and licensing of big data.
- The similarities and differences between big and long tail data requirements, when it comes to the design of a VRE, could be investigated.
- An investigation could be done on how a formal structure for RDM could be established at a higher education institution, and if this could be generalised for use in other universities. This investigation could look at things such as RDM strategy, institutional culture, IT infrastructure, policy framework, ethical processes, research information systems, etc.
- An investigation on possible criteria for the evaluation of different data repository platforms could be conducted. Such an investigation could look at functional criteria, for example, deposit and upload, re-usability, identity and access management, reporting, and preservation. Non-functional criteria could also be investigated, for example, back-end management, integration, infrastructure, vendor specific criteria, level of training needed, and ease of use.

8.8 CONCLUDING REMARKS

This study has shown that VREs as technology frameworks can facilitate research projects at a university. It also showed that VREs join together a number of components (human, hardware, software, management, standards, protocols and specifications, and policy components) that interact with one another and could be utilised in the research lifecycle of a research project.

The conceptual framework model with its various layers and components, as proposed in Chapter 3 under 3.5.7 and further enhanced in Chapter 5 under 5.4, was verified through the empirical study as discussed under 8.2.3. The empirical study, however, revealed that a small number of components could be omitted; however, it also contributed a number of new components to the model. The researcher of this study then refined the conceptual model in 8.2.7 and in Figures 8.1a-d, to accommodate these changes. The empirical study showed that the VRE conceptual model as presented in 8.2.7 could potentially be used in other environments, but because the study was only limited to two case studies, the conceptual model could not be generalised to multiple disciplinary areas/environments. The discussion in 8.6, nonetheless, revealed that guidelines for the use of the VRE conceptual model that was proposed in 8.2.7, could be applied in multiple disciplinary areas / environments.

The study also disclosed that the success of the two VRE case studies was very much dependent on the user-friendliness and intuitiveness of the software platforms used, the enthusiasm and encouragement of the VRE Managers and VRE Champion, the individual training received, as well as the communication between members and communication between members and the VRE Designer.

An important issue that arose during the study was that a VRE's interface has to be integrated with a researcher's everyday activities and research workflow, and not be seen as an add-on. In other words, it should make life easier for a researcher (see 7.3.1.1 under Question 5). This links up with the criticism that was raised against the Alfresco VREs, in that they were seen as separate from existing work processes and workflows of these researchers, and also not intuitive enough (see 7.3.1.1 under Question 23 (c)). This could also be one of the reasons why some of the components in the VREs were not used, and why the VREs in the case studies were mostly used for the storage and back-up of data.

The major outcome of this study was that VREs were revealed as ideal instruments that could be used in the successful management of research data (see 8.2.6). In fact, the case studies showed that the majority of respondents used the VREs mainly for the management of their data. RDM and all its components were also shown as an essential part of the functioning of a VRE and its purpose to support the research lifecycle, as well as the research data lifecycle, consisting of the capturing, processing, analysis, preservation, sharing and re-use of researchers' data.

BIBLIOGRAPHY

About MyTardis, n.d. [Sl.: sn.] [Online] available at http://www.mytardis.org/about/ (Accessed 8 February 2017).

Academia.edu, n.d. San Francisco, CA: Academia.edu. [Online] available at http://www.academia.edu (Accessed 24 March 2013).

ACKOFF, R.L. 1989. From data to wisdom. ***Journal of Applied Systems Analysis***, 16(1): 3-9.

ACLS. 2006. ***Our cultural commonwealth: the report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences.*** New York, NY: American Council of Learned Societies (ACLS). [Online] available at http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf (Accessed 21 February 2015).

Africa Centre for population health, n.d. [Online] available at http://www.africacentre.ac.za (Accessed 5 January 2017).

ALAMEDA, J., CHRISTIE, M., FOX, G., FUTRELLE, J., GANNON, D., HATEGAN, M., KANDASWAMY, G., VON LASZEWSKI, G., NACAR, M.A., PIERCE, M., ROBERTS, E., SEVERANCE, C. & THOMAS, M. 2007. The Open Grid Computing Environments collaboration: portlets and services for science gateways. Concurency and Computation: Practice and Experience, 19: 921-942.

Alfresco. [Maidenhead, UK]: Alfresco Software, Inc, 2012. [Online] available at http://www.alfresco.com/ (Accessed 9 August 2012).

ALLAN, R. 2009. ***Virtual Research Environments: from portals to science gateways.*** Oxford: Chandos Publishing.

ALLIANCE OF GERMAN SCIENCE ORGANISATIONS. 2008. ***Priority Initiative 'Digital Information'.*** [Sl.]: Alliance of German Science Organisations. [Online] available at http://www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/AllianInitiative_englisch.pdf (Accessed 21 September 2017).

ALLIANCE OF GERMAN SCIENCE ORGANISATIONS. 2013. ***Priority Initiative 'Digital Information': extending the cooperation 2013-2017***. [Sl.]: Alliance of German Science Organisations. [Online] available at http://www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/Priority_Initiative_2013-2017.pdf (Accessed 21 September 2017).

ALMATURI, A.F., GARDNER, G.E. & MCCARTHY, A. 2014. Practical guidance for the use of a pattern-matching technique in case-study research: a case presentation. ***Nursing and Health Sciences***, June, 16(2): 239-244.

ALTINTAS, I., BARNEY, O., CHENG, Z., CRITCHLOW, T., LUDAESCHER, B., PARKER, S., SHOSHANI, A. & VOUK, M. 2006. Accelerating the scientific exploration process with scientific workflows. ***Journal of Physics: Conference Series***, 46: 468-478. [Online] available at http://iopscience.iop.org/article/10.1088/1742-6596/46/1/065/pdf (Accessed 12 February 2018).

ALTMAN, M. & KING, G. 2007. A proposed standard for the scholarly citation of quantitative data. ***D-Lib Magazine***, March / April, 13(3/4). [Online] available at http://www.dlib.org/dlib/march07/altman/03altman.html (Accessed 12 October 2014).

ALZFORUM. 2017. ***Databases.*** Cambridge, MA: AlzForum. [Online] available at http://www.alzforum.org/databases (Accessed 21 September 2017).

AMERICAN PSYCHOLOGICAL ASSOCIATION. 2009. *Publication Manual of the American Psychological Association.* WASHINGTON, DC: American Psychological Association.

Analyzing and interpreting data. Syracuse, NY: Office of Institutional Research and Assessment, Syracuse University, n.d. [Online] available at https://oira.syr.edu /assessment/assesspp/Analyze.htm (Accessed 18 September 2014).

ANDERSON, S., DUNN, S. & HUGHES, L.M. 2005. VREs in the arts and humanities. In ***Proceedings of the UK e-Science All Hands Meeting 2005, 19-22 September, Nottingham, UK.*** Edited by Simon J. Cox and David W. Walker. [Swindon, UK]: EPSRC, p. 514-517. [Online] available at http://www.allhands.org.uk/2005/ proceedings/proceedings/proceedings.pdf (Accessed 1 May 2015).

ARC: African Research Cloud. Cape Town: University of Cape Town, 2017. [Online] available at http://www.arc.ac.za/ (Accessed on 12 January 2017).

Archer project. [Sl.: s.n.], n.d. [Online] available at http://archer.edu.au/ (Accessed 25 September 2012).

ARL. 2007. ***Agenda for Developing E-Science in Research Libraries.*** Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee. Prepared by the Joint Task Force on Library Support for E-Science. Washington, DC: Association of Research Libraries. [Online] available at http://www.arl.org/bm~doc/ARL_EScience_final.pdf (Accessed 26 February 2012).

ASTRON AND IBM CENTER FOR EXASCALE TECHNOLOGY. 2017. ***Dome Simposium: New Foundations for a Smart Society, 18-19 May 2017, Dwingeloo, Netherlands.*** [Online] available at http://www.dome-exascale.nl/symposium2017/ (Accessed 13 September 2017).

ATKINS, D.E., DROEGEMEIER, K.K., FELDMAN, S.I., GARCIA-MOLINA, H., KLEIN, M.L., MESSERSCHMITT, D.G., MESSINA, P., OSTRIKER, J.P. & WRIGHT, M.H. 2003. ***Revolutionizing science and engineering through cyberinfrastructure.*** Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. [Arlington, VA: National Science Foundation]. Online available at http://www.nsf.gov/cise/sci/reports/atkins.pdf (Accessed 23 February 2015).

AUSTRALIAN GOVERNMENT. 2015. ***Australian Government Public Data Policy Statement.*** [Canberra, ACT]; Australian government. [Online] available at: https://www.pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf (Accessed 12 September 2017).

AUSTRALIAN GOVERNMENT, AUSTRALIAN RESEARCH COUNCIL. 2015. ***Research data management.*** Canberra, ACT: Australian Government, Australian Research Council. [Online] available at http://www.arc.gov.au/research-data-management (Accessed 12 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF EDUCATION AND TRAINING. 2015. ***Funded research infrastructure projects.*** [Canberra, ACT]: Australian Government, Department of Education and Training. [Online] available at https://www.education.gov.au/funded-research-infrastructure-projects (Accessed 9 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF EDUCATION AND TRAINING. 2016. ***National Research Infrastructure Roadmap.*** [Canberra, ACT]: Australian Government, Department of Education and Training. [Online] available at https://docs.education.gov.au/system/files/doc/other/ed16-0269_national_research_infrastructure_roadmap_report_internals_acc.pdf (Accessed 9 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF EDUCATION AND TRAINING. 2017. ***National Collaborative Research Infrastructure Strategy (NCRIS).*** [Canberra, ACT]: Australian Government, Department of Education and Training. [Online] available at https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris (Accessed 9 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF EDUCATION, SCIENCE AND TRAINING. 2004. ***National Research Infrastructure Framework: the final report of the National Research Infrastructure Taskforce.*** Canberra, ACT: Australian Government, Department of Education, Science and Training. [Online] available at https://docs.education.gov.au/system/files/doc/other/national_research_infrastructure_taskforce_final_report_2004.pdf (Accessed 8 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF EDUCATION, SCIENCE AND TRAINING. 2006. ***National Collaborative Research Infrastructure Strategy Strategic Roadmap.*** Canberra, ACT: Australian Government, Department of Education. [Online] available at: http://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2006.pdf (Accessed 8 September 2017).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF INNOVATION, INDUSTRY, SCIENCE AND RESEARCH. 2008. ***Strategic Roadmap for Australian Research Infrastructure.*** Canberra, ACT: Australian Government, Department of Innovation, Industry, Science and Research. [Online] available at https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2008.pdf (Accessed 9 September 2017).

AUSTRALIAN GOVERNMENT, NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL. 2014. ***Principles for accessing and using publicly-funded data for health research: targeted consultation draft.*** [Canberra, ACT]: Australian Government, National Health and Medical Research Council. [Online] available at https://consultations.nhmrc.gov.au/files/consultations/drafts/draftprinciplesaccessingpubliclyfundeddata141209.pdf (Accessed 12 September 2017).

AUSTRALIAN NATIONAL DATA SERVICE. n.d.(a) ***What we do***. [Caulfield East, VIC and Acton, ACT: Australian National Data Service]. [Online] available at http://www.ands.org.au/about-us/what-we-do (Accessed 4 January 2017).

AUSTRALIAN NATIONAL DATA SERVICE. n.d.(b) ***Research Data Australia***. [Caulfield East, VIC and Acton, ACT: Australian National Data Service]. [Online] available at http://www.ands.org.au/online-services/research-data-australia (Accessed 12 September 2017).

AUSTRALIAN NATIONAL DATA SERVICE. n.d.(c) ***Completed programs***. [Caulfield East, VIC and Acton, ACT: Australian National Data Service]. [Online] available at http://www.ands.org.au/partners-and-communities/projects/completed-programs (Accessed 12 September 2017).

AUSTRALIAN NATIONAL DATA SERVICE. 2011. ***Research data management framework: capability maturity guide.*** [Online] available at: https://web.archive.org/web/20150921052702/http://ands.org.au/guides/dmframework/dmf-capability-maturity-guide.pdf (Accessed 9 September 2017).

AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES. n.d. **About.** [Canberra, ACT]: Australian Partnership for Sustainable Repositories. [Online] available at: http://apsr.anu.edu.au/about.html (Accessed 9 September 2017).

AUSTRALIAN RESEARCH COUNCIL. 2005. ***ARC e-Research Support: invitation for funding proposals under ARC Special Research Initiatives for funding to commence in 2005.*** [Canberra, ACT]: Australian Government, Australian Research Council. [Online] available at http://www.arc.gov.au/pdf/Invitation_for_Funding_Proposals_ER05_060105.pdf (Accessed 18 September 2011).

AUSTRALIAN RESEARCH COUNCIL. 2014. ***Funding rules for schemes under the Discovery Program for the years 2014 and 2015.*** [Canberra, ACT]: Australian Government, Australian Research Council. [Online] available at http://archive.arc.gov.au/archive_files/Funded%20Research/1%20Discovery%20Program/Discovery%20Projects/2015/DP15_Funding_Rules.pdf (Accessed 9 September 2017).

AVGEROU, C. 2000. Information systems: what sort of science is it? ***Omega***, 1 October, 28(5): 567-579.

BABBIE, E. & MOUTON, J. 2001.***The practice of social research.*** Oxford, UK: Oxford University Press.

BAI, G. 2014. An organic view of prototyping in information system development. In: ***Proceedings of the IEEE 17th International Conference on Computational Science and Engineering, 19-21 December, Chengdu, China.*** Piscataway, NJ: The Institute of Electrical and Electronics Engineers (IEEE), p. 1814-1818.

BALL, A. 2012. ***Review of Data Management Lifecycle Models.*** (version 1.0). REDm-MED Project Document, redm1rep120110ab10. Bath, UK: University of Bath. [Online] available at http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf (Accessed 30 September 2014).

BARGA, R.S., ANDREWS, S. & PARASTATIDES, S. 2007. A Virtual Research Environment (VRE) for Bioscience Researchers. In: ***Proceedings of the International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP)***, ***Papeete, 4-9 November, French Polynesia.*** Washington, DC: IEEE, p.31-38.

BARNES, T.A., PASHBY, I.R. & GIBBONS, A.M. 2006. Managing collaborative R&D projects development of a practical management tool. ***International Journal of Project Management***, 24(5): 395-404.

BASNEY, J., DOOLEY, R., GAYNOR, J., MARRU, S. & PIERCE, M. 2011. Distributed web security for Science Gateways. In: ***GCE '11, Proceedings of the 2011 ACM workshop on Gateway computing environments, 18 November, Seattle, Washington.*** New York, NY: ACM, p.13-20.

BEAGRIE, N. 2004. The continuing access and digital preservation strategy for the UK Joint Information Systems Committee (JISC). ***D-Lib Magazine***, July / August, 10(7/8). [Online] available at http://www.dlib.org/dlib/july04/beagrie/07beagrie.html (Accessed 30 September 2014).

BEAULIEU, A. & WOUTERS, P. 2009. E-research as intervention. In: ***E-Research: transformation in scholarly practice.*** Edited by Nicholas W. Jankowski. New York, NY: Routledge, p. 54-69.

BEITZ, A., DHARMAWARDENA, K. & SEARLE, S. 2012. ***Monash University: research data management strategy and strategic plan 2012-2015.*** Version for public release 13, April 2012. [Melbourne, Australia]: Monash University. https://confluence.apps.monash.edu/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-publicrelease.pdf?version=1&modificationDate=1334289180000 (Accessed 3 January 2017).

BELL, G. 2009. Foreword. In: ***Fourth paradigm: data-intensive scientific discovery.*** Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Redmond, WA: Microsoft Research, c2009, p. xv.

BENOIT, K. 2011. Data, textual. In: ***International Encyclopedia of Political Science.*** Edited by Bertrand Badie, Dirk Berg-Schlosser and Leonardo Morlino. Thousand Oaks, CA: Sage Publications, p. 526-531.

BERG, B.L. 1995. ***Qualitative research methods for the social sciences.*** 4th ed. Boston, MA: Allyn and Bacon.

BERMAN, F. & BRADY, H. 2005. ***Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences, 12 May 2005.*** North-Arlington, VA: National Science Foundation. [Online] available at http://ucdata.berkeley.edu/pubs/CyberInfrastructure_FINAL.pdf (Accessed 21 February 2015).

BERNARD BECKER MEDICAL LIBRARY. 2018. Scholarly communications at Becker. St. Louis, MO: Bernard Becker Medical Library, Washington University in St. Louis. [Online] available at https://becker.wustl.edu/expertise/scholarly-communications (Accessed 12 February 2018).

BERTRAND, I. & HUGHES, P. 2005. ***Media research methods: audiences, institutions, texts.*** Basingstoke: Palgrave Macmillan.

BEZUIDENHOUT, R. 2016. ***The research data management journey at UNISA.*** Presented at the NeDiCC meeting, CSIR, Pretoria, 18 February 2016. [Unpublished].

BigBlueButton. Ottawa, ON: BigBlueButton Inc., 2012. [Online] available at http://www.bigbluebutton.org/ (Accessed 8 May 2013).

BIGELOW, S.J. & ROUSE, M. 2016. Platform. ***WhatIs.com***, web log post, 9 August 2016. [Online] available at http://searchservervirtualization.techtarget.com/definition /platform (Accessed 12 February 2018).

BISCHOFBERGER, W. & POMBERGER, G. 1992. ***Prototyping-oriented software development: concepts and tools.*** (Texts and Monographs in Computer Science). New York, NY: Springer.

Blackboard. Washington, D.C.: Blackboard Inc, 2012. [Online] available at http://www.blackboard.com (Accessed on 28 September 2012).

BLOMKVIST, J. 2014. ***Representing future situations of service: prototyping in service design.*** (Linköping Studies in Arts and Science, No. 618). Linkoping, Sweden: Linkoping University. [Online] available at https://liu.diva-portal.org/smash/get/ diva2:712357/FULLTEXT02.pdf (Accessed 17 January 2017).

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

BOS, N., ZIMMERMAN, A., OLSON, J., YEW, J., YERKIE, J., DAHL, E. & OLSON, G. 2007. From shared databases to Communities of Practice: a taxonomy of collaboratories. ***Journal of Computer-Mediated Communication***, January, 12(2): 652-672.

BOTHMA, T.J.D., PIENAAR, H. & HAMMES, M. 2008. Towards open scholarship at the University of Pretoria. In: ***16 th BOBCATSSS Symposium 2008: providing access to information for everyone: Zadar, Croatia, 28-30 January 2008: proceedings.*** Bad Honnef, Germany: Bock & Herchen Verlag, p. 271 – 284. [Online] Available /edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=28487 (Accessed 11 February 2018).

BOUTELLIER, R., GASSMANN, O., RAEDER, S., DÖNMEZ, D. & DOMIGALL, Y. 2011. What is the difference between social and natural sciences? Presented at ***Doctoral Seminar "Forschungsmetodik I" HS11-10, 118, 1.00***, Fall Semester, Universität St Gallen, Zurich, Switzerland. [Online] available at http://www.collier.sts.vt.edu/sciwrite/pdfs/boutellier_2011.pdf (Accessed 16 February 2017).

BOWERS, N. & VAN DEVENTER, M. 2012. Virtual research environments: the role of the facilitator. Presented at ***11th SAOIM Meeting***, 7 June, Johannesburg Convention Centre, Johannesburg. [Online] available at http://www.saoug.files.wordpress.com/2012/06/natalie1.pptx *(Accessed 3 October 2012).*

BRADLEY, L. 2013. ANU Library's support for research data management. ***Incite***, April, 34(4): 26.

BROWN, C. 2012. ***JISC's VRE programme: supporting collaborative research.*** [PPT Presentation] [Online] available at http://www.slideshare.net/chriscb/jisc-vreresearch-tools-presentation (Accessed 30 June 2012).

BROWN, C. 2013. **VRE and research tools: supporting collaborative research.** Presented at the ***JISC Regional Support Centre London Webinar,*** 20 November London, UK. [Online] available at http://www.slideshare.net/chriscb/vres-and-research-tools-supporting-collaborative-research?related=1 (Accessed 5 February 2015).

BROWN, C. 2017. ***Next Generation Research Environments: recommendations and next steps.*** [Sl.]: Joint Information Systems Committee (JISC). [Online] available at https://researchdata.jiscinvolve.org/wp/2017/07/19/ngre-recommendations/ (Accessed 21 September 2017).

BROWN, S., BRUCE, R. & KERNOHAN, D. 2015. ***Directions for research data management in UK Universities.*** [Sl.]: JISC. [Online] available at http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf (Accessed 2 January 2017).

BRYMAN, A. 2001. ***Social research methods.*** Oxford; New York: Oxford University Press.

Business dictionary, 2018. Austin, TX: WebFinance. [Online] available at http://www.businessdictionary.com/definition/component.html (Accessed 12 February 2018).

Business logic. ***WhatIs.com,*** 2015***.*** [Online] available at http://whatis.techtarget.com/definition/business-logic (Accessed 21 February 2015).

CALIFORNIA DIGITAL LIBRARY. 2014. ***DMP Tool: guidance and resources for your data management plan.*** Oakland, CA: University of California Curation Center of the California Digital Library. [Online] available at https://dmp.cdlib.org/ (Accessed 13 March 2014).

CALLAGHAN, S., MURPHY, F., TEDDS, J., ALLAN, R., KUNZE, J., LAWRENCE, R., MAYERNIK, M.S., WHYTE, A. & PREPARDE PROJECT TEAM. 2013. Connecting data repositories and publishers for data publication. Presentation delivered on 7 February 2013 at the ***OpenAIRE Interoperability Workshop*****,** 7-8 February, University of Minho Gualtar Campus, Braga, Portugal. [Online] available at http://openaccess.sdum.uminho.pt/wp-content/uploads/2013/02/7_SarahCallaghan_OpenAIREworkshopUMinho.pdf (Accessed 19 September 2014).

CANDELA, L. n.d. ***GRDI2020: Virtual Research Environments.*** [Sl.: s.n.] [Online] available at http://www.grdi2020.eu/Repository/FileScaricati/eb0e8fea-c496-45b7-a0c5-831b90fe0045.pdf (Accessed 1 March 2015).

CANDELA, L., CASTELLI, D. & PAGANO, P. 2009. On-demand virtual research environments and the changing roles of librarians. ***Library Hi Tech***, 27(2): 239-251.

CANDELA, L., CASTELLI, D. & PAGANO, P. 2010. Making Virtual Research Environments in the cloud a reality: the gCube approach. ***ERCIM News*****,** Online Edition, October, 83: 32-33. [Online] available at http://ercim-news.ercim.eu/Images/stories/EN83/EN83-web.pdf (Accessed 20 January 2013).

CAPURRO, R. 1978. ***Information: Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs*** [A contribution to the etymological and conceptual history of the concept of information]. München: Saur Verlag.

CARUSI, A. & REIMER, T. 2010. ***Virtual Research Environment collaborative landscape study: a JISC funded project.*** [Bristol, UK: JISC]. [Online] available at http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf (Accessed 26 February 2012).

CAUDET, L., VANDYSTADT, N., FOUGNER, A. & FRENAY, M. 2016. ***European Cloud Initiative to give Europe a global lead in the data-driven economy: European Commission Press release.*** (IP/16/1408) [Sl.: European Commission]. [Online] available at http://europa.eu/rapid/press-release_IP-16-1408_en.htm (Accessed 1 February 2017).

CESARSKY, C., DONAHUE, M., FARRARESE, L., MURONGA, A., SMITH, C. & WOUDT, P. 2015. ***Southern African Large Telescope five-year review.*** [Sl.: s.n) [Online] available at http://www.salt.ac.za/wp-content/uploads/sites/75/2017/06/3.10-SALT-Review-2016_Final.pdf (Accessed 17 September 2017).

CHAIN-REDS. n.d. ***Science Gateway.*** [Sl.]: Co-ordination and Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing (CHAIN-REDS). [Online] available at http://science-gateway.chain-project.eu/ (Accessed 12 February 2015).

CHAMBERS, S. 2002. Supporting teaching and research in an online environment: developing the University of London Library model. ***LIBER Quarterly***, 12(4): 381-392. [Online] available at http://liber.library.uu.nl/index.php/lq/article/view/7704/7740 (Accessed 12 August 2012).

CHARLES, K. 2012. ***Comparing enterprise data anonymization techniques.*** Newton, MA: TechTarget. [Online] available at http://searchsecurity.techtarget.com/tip/Comparing-enterprise-data-anonymization-techniques (Accessed 18 September 2014).

Chisimba.com. [Sl.: AVOIR], n.d. [Online] available at http://chisimba.com/ (Accessed 4 October 2012).

CHIWARE, E. & MATHE, Z. 2015. Academic Libraries' role in research data management services: A South African perspective. ***South African Journal of Libraries and Information Science***, 81(2): 1-10. [Online] available at http://sajlis.journals.ac.za/pub/article/view/1563 (Accessed 16 September 2017).

CHOHAN, D. 2005. CCLRC portal infrastructure to support research facilities. Presented at the ***Science Gateway Workshop GGF14***, Chicago, Illinois, 28 June 2005. [Online] available at http://studyslide.com/doc/556362/cclrc-portal-infrastructure-to-support-research (Accessed 10 February 2018).

CHOUDHURY, S. 2013. The research data revolution. Presented at the ***STM Innovations Seminar***, 4 December, Congress Centre, London, UK. [Online] available at http://www.stm-assoc.org/2013_12_04_Innovations_Choudhury_The_Research_Data.pdf (Accessed 28 August 2014).

CHPC: CENTRE FOR HIGH PERFORMANCE COMPUTING. n.d. ***About us.*** [Cape Town]: CHPC: Centre for High Performance Computing [Online] available at http://www.chpc.ac.za/index.php/about-us/mission-objectives (Accessed 7 July 2014).

CHRISTENSSON, P. 2008. Application definition. ***TechTerms.*** [Online] available at https://techterms.com/definition/application (Accessed 12 February 2018).

CHRISTENSSON, P. 2009. User interface definition. ***TechTerms.*** [Online] available at https://techterms.com/definition/user_interface (Accesed 12 February 2018).

Circular 6/03 (Revised) Digital Curation Centre. [Sl.: JISC], 2012. [Online] available at https://archive.fo/Rhwg#selection-329.3-329.51 (Accessed 4 December 2016).

CODATA. n.d. ***African Open Science Plaform to boost the impact of open data for science and society: media release, 8 December 2016.*** Paris, France: CODATA, International Council for Science: Committee on Data for Science and Technology. [Online] available at http://www.codata.org/news/150/62/African-Open-Science-Platform-to-boost-the-impact-of-open-data-for-science-and-society-Media-Release (Accessed 19 September 2017).

CODATA. 2017. ***Who are we?*** [Paris, France?]: CODATA: Committee on Data for Science and Technology, International Council for Science and Technology. [Online] available at http://www.codata.info/about/who.html (Accessed 23 September 2017).

CODATA Workshop on Archiving Scientific & Technical (S&T) DATA, 20-21 May 2002, Pretoria, South Africa: report. Pretoria: South African National Committee for CODATA; CODATA Working Group on Data Archiving; National Research Foundation, 2002. [Online] available at http://stardata.nrf.ac.za/COdata/CodataReport_2002.pdf (Accessed 19 August 2014).

COGBURN, D.L. 2003. HCI in the so-called developing world: what's in it for everyone? ***Interactions***, March and April, 10(2): 80-87.

THE COLLEGE AT BROCKPORT. 2012. ***Computational science: what is computational science?*** Brockport, NY: The College at Brockport, State University of New York. [Online] available at http://www.brockport.edu/cps/whatis.html (Accessed 20 January 2012).

COLLIN, S.M.H. 2002. ***Dictionary of Computing.*** 4th ed. London, UK: Peter Collin Publishing, p. 1-532. [Online] available at https://epdf.tips/dictionary-of-computing-over-10000-terms-clearly-defined.html (Accessed 12 February 2018).

COLUMBIA UNIVERSITY LIBRARIES. n.d. ***Why manage research data?*** New York, NY: Columbia University Libraries. [Online] available at http://scholcomm.columbia.edu/data-management/why/ (Accessed 21 October 2014).

COMMISSION HIGH LEVEL EXPERT GROUP ON THE EUROPEAN OPEN SCIENCE CLOUD. 2016. ***Realising the European Open Science Cloud: first report and recommendations.*** Brussels: European Commission, Directorate General for Research and Innovation. [Online] available at https://ec.europa.eu/research/Openscience/pdf/realising_the_european_open_science_cloud_2016.pdf (Accessed 19 September 2017).

COMMISSIONERATE OF COLLEGIATE EDUCATION. 2017. ***What is e-Learning?*** Prasadmapady, Vijayawada, India: Commissionerate of Collegiate Education, Government of Andhra Pradesh, India. [Online] available at http://www.apcce.gov.in/newwebsite30122010/whatiselearn.aspx (Accessed 10 February 2018).

CONCORDAT WORKING GROUP. 2016. ***Concordat on Open Research Data.*** [Sl.]: Concordat Working Group. [Online] available at http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf (Accessed 30 August 2017).

CONCORDIA UNIVERSITY LIBRARIES. 2014 (Change to 2017). **Why manage your data?** Montreal, QC: Concordia University. [Online] available at https://library.concordia.ca/help/data/why-data.php (Accessed 24 September 2017).

CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES (CESSDA). 2016. ***Mandate.*** [Sl.]: Consortium of European Social Science Data Archives (CESSDA). [Online] available at http://cessda.net/About-us/Mandate (Accessed 23 May 2016).

CORMODE, G. & SRIVASTAVA, D. 2009. Anonymized data: generation, models, usage. Tutorial presented at the ***2009 ACM SIGMOD International Conference on Management of Data***, 2 July, Providence, Rhode Island, New York. [Online] available at http://dimacs.rutgers.edu/~graham/pubs/papers/anontut.pdf (Accessed 17 September 2014).

CORTI, L., VAN DEN EYNDEN, V., BISHOP, L. & WOOLLARD, M. 2014. ***Managing and sharing research data: a guide to good practice.*** Los Angeles: SAGE.

COX, A.M., KENNAN, M.A., LYON, L. & PINFIELD, S. 2017. Developments in research data management in academic libraries: towards an understanding of research data service maturity. ***Journal of the Association for Information Science and Technology***, March, 68(9): 2182-2200. [Online] available at http://onlinelibrary.wiley.com/doi/10.1002/asi.23781/epdf (Accessed 4 September 2017).

CRAGIN, M.H., HEIDORN, P.B., PALMER, C.L. & SMITH, L.C. 2007. An educational program on data curation. Poster presented at the ***2007 STS Conference Poster Session,*** 25 June, Washington, DC. [Online] available at http://hdl.handle.net/2142/3493 (Accessed 18 August 2014).

CRAWFORD, S. 2013. Big data from small telescopes. ***Steve Crawford, SALT Science Data Manager: Research Blog.*** Posted 29 October 2013. [Online] available at http://stevecrawford.saao.ac.za/2013/10/29/big-data-from-small-telescopes/ (Accessed 28 April 2014).

Creative Commons. Mountain View, CA: Creative Commons, n.d. [Online] available at http://www.creativecommons.org (Accessed 18 January 2017).

CRESWELL, J.W. 2003. ***Research design: qualitative, quantitative and mixed method approaches.*** Thousand Oaks: Sage.

CRESWELL, J.W. 2007. ***Qualitative enquiry and research design: choosing among five approaches.*** 2nd ed. Thousand Oaks: CA: Sage.

CREW, R.M. 2007. ***Policy for the preservation and retention of research data***. (Rt306/07). Pretoria: University of Pretoria.

CROTTY, M. 1998. ***The foundations of social research: meaning and perspective in the research process.*** London, UK: Sage.

CSERD. 2012. ***What is computational science?*** [Sl.]: The Shodor Education Foundation, Computational Science Education Reference Desk (CSERD). [Online] available at http://www.shodor.org/csedr/Help/whatiscs (Accessed 20 January 2012).

CSIR. 2011. ***Overview: Cyberinfrastructure***, 2011. [Online] available at http://www.csir.co.za/meraka/cyberinfrastructure/ (Accessed 7 July 2014).

CSIR. 2017. ***Explore the CSIR.*** [Pretoria?]: CSIR. [Online] available at https://wwwprod.csir.co.za/about-us (Accessed 6 January 2017).

Data Curation Profiles Toolkit, n.d. [Sl.: S.n.]. [Online] available at: http://datacurationprofiles.org/ (Accessed 13 March 2014).

DAMA Dictionary of Data Management. 2nd.ed. Bradley Beach, N.J.: Technics Publications, 2011.

DAMA INTERNATIONAL. 2007. ***DAMA-DMBOK Guide: Data Management Body of Knowledge: Introduction and Project Status, November 2007***. Edited by Mark Mosley. Middletown, DE: DAMA International. [Online] available at http://www.dama.org/files/public/DI_DAMA_DMBOK_Guide_Presentation_2007.pdf (Accessed 29 August 2014).

DAMA INTERNATIONAL. 2014. ***About us.*** Middletown, DE: DAMA International. [Online] available at http://www.dama.org/i4a/pages/index.cfm?pageid=3339 (Accessed 19 August 2014)

DARRIES, F. 2016. ***Introducing research data management at Unisa: a small presentation on small beginnings.*** Presentation at the NeDiCC Meeting, 14 September, CSIR, Pretoria, 14 September 2016. [Unpublished]

Data Carpentry, 2017. [Sl.: s.n.] [Online] available at http://www.datacarpentry.org/ (Accessed 17 September 2017).

DataCite, n.d. [Sl.: s.n.]. [Online] available at http://www.datacite.org/ (Accessed 30 July 2014).

DataConservancy, n.d. Baltimore, MD: Johns Hopkins University, The Sheridan Libraries, Data Conservancy. [Online] available at http://dataconservancy.org/ (Accessed 2 September 2017).

Data Curation Profiles Toolkit*, n.d.* [Sl.: s.n.]. [Online] available at http://datacurationprofiles.org/ (Accessed 25 October 2014).

DATAFIRST. 2017. ***About DataFirst.*** Cape Town: DataFirst, University of Cape Town. [Online] available at https://www.datafirst.uct.ac.za/about-us (Accessed 17 September 2017).

Data Management for NSF SBE Directorate: proposals and awards. Arlington, VA: National Science Foundation, n.d. [Online] available at http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf (Accessed 25 October 2014).

DataNet Federation Consortium, 2017. [Online] available at http://datafed.org/ (Accessed 2 September 2017).

DataONE and USGS: making open data a reality, 2016. ***DataONE News***, Spring, 4(3): 1-7. [Online] available at https://www.dataone.org/sites/default/files/sites/all/documents/newsletters/dataonenews_spring2016v2_sm.pdf (Accessed 3 September 2017).

DataONE: Data Observation Network for Earth, n.d. Albuquerque, NM: University of New Mexico Albuquerque. [Online] available at http://www.dataone.org/ (Accessed on 13 March 2014).

Data Seal of Approval, n.d. [Sl.: s.n.] [Online] available at http://www.datasealofapproval.org/en/ (Accessed 25 November 2016).

DCC. 2014a. ***Overview of funders data policies.*** Edinburgh, UK: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies (Accessed 24 September 2014).

DCC. 2014b. ***DCC Curation Lifecycle Model.*** Edinburgh, UK: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/resources/curation-lifecycle-model (Accessed 30 September 2014).

DCC. 2014c. ***DCC Charter and Statement of Principles.*** Edinburgh, UK: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/about-us/dcc-charter/dcc-charter-and-statement-principles (Accessed 18 August 2014).

DCC. 2014d. ***DMPonline.*** Edinburgh, UK: Digital Curation Centre, 2014. [Online] available at https://dmponline.dcc.ac.uk/ (Accessed 29 September 2014).

DCC. 2016. ***History of the DCC.*** Edinburgh, UK: DCC. [Online] available at http://www.dcc.ac.uk/about-us/history-dcc/history-dcc (Accessed 3 December 2016).

DCC. 2017. ***Data Asset Framework.*** Edinburgh, UK: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/resources/tools/data-asset-framework (Accessed 30 August 2017).

DE LA FLOR, G. & MEYER, E.T. 2008. *Talking 'bout a revolution: framing e-Research as a computerization movement.* Presented at the Oxford eResearch Conference, 11-13 September, Oxford Internet Institute and Oxford e-Research Centre, University of Oxford, Oxford. [Online] available at https://www.slideshare.net/etmeyer/talking-bout-a-revolution-framing-eresearch-as-a-computerization-movement-presentation (Accessed 24 September 2017).

Delivering research data services: fundamentals of good practice. Edited by Graham Pryor, Sarah Jones and Angus Whyte. London: Facet Publishing, 2014.

DEMPSEY, L. 2007. Data curation education. ***Lorcan Dempsey's Weblog***, web log post, 25 February, 2007. [Online] available at http://orweblog.oclc.org/archives/001277.html (Accessed 17 August 2014).

DE ROURE, D., GOBLE, C., BHAGAT, J., CRUICKSHANK, D., GODERIS, A., MICHAELIDES, D. & NEWMAN, D. 2008. MyExperiment: defining the social Virtual Research Environment. *In:* ***Proceedings of the 4th IEEE International Conference on e-Science, 7-12 December 2008, Indianapolis, Indiana.*** [Sl.]: IEEE Press, p. 182-189. [Online] available at http://eprints.ecs.soton.ac.uk/16560 (Accessed 6 January 2013).

DE ROURE, D., GOBLE, C., ALEKSEJEVS, S., BECHHOFER, S., BHAGAT, J., CRUIKSHANK, D., MICHAELIDES, D. & NEWMAN, D. 2009. The myExperiment open repository for scientific workflows. Paper presented at the ***4th International Conference on Open Repositories 2009,*** 18-21 May, Atlanta, Georgia. [Online] available at http://eprints.soton.ac.uk/267131/ (Accessed 6 January 2013).

DE WALT, K.M., DE WALT, B.R. & WAYLAND, C.B. 1998. Participant observation. In: ***Handbook of methods in Cultural Anthropology***. Edited by H. Russell Bernard. Walnut Creek, CA: Altamira Press, p.259-299.

DIAMOND, C.C., MOSTASHARI, F. & SHIRKY, C. 2009. Collecting and sharing data for population health. ***Health Affairs***, March / April, 28(2): 454-466.

DIETRICH, D., ADAMUS, T., MINER, A. & STEINHART, G. 2012. De-mystifying the data management requirements of research funders. ***Issues in Science and Technology Librarianship***, Summer, 70. [Online] available at http://www.istl.org/12-summer/refereed1.html (Accessed 22 September 2012).

Digital Collaboratory for Cultural Dendrochronology (DCCD): an international digital data library for dendrochronology, n.d. [Online] available at http://dendro.dans.knaw.nl/ (Accessed 20 April 2013).

DILLENBOURG, P. 2000. Virtual Learning Environments. Presented at the ***Workshop on Virtual Learning Environments, EUN Conference 2000: Learning in the New Millennium: Building New Education Strategies for Schools, University of Geneva, 20-21 March, Geneva, Switzerland.*** [Online] available at: https://tecfa.unige.ch/tecfa/publicat/dil-papers-2/Dil.7.5.18.pdf (Accessed 12 February 2018).

DI MURO, D. & SAUNDERS, E. 2008. Virtual research environments: what do libraries have to do with it? A paper presented at ***ALIA Web2.0: Beyond the Hype Symposium***, 1-2 February, Gardens Point Campus, Queensland University of Technology, Brisbane. [Online] available at http://unsworks.unsw.edu.au/fapi/datastream/unsworks:181/SOURCE01 (Accessed 12 August 2012).

DIRISA. 2017. ***Figshare pilot in Pretoria and Durban***. Pretoria: DIRISA. [Online] available at https://www.dirisa.ac.za/figshare-pilot-pretoria-durban/ (Accessed 16 September 2017). DIGITAL LIBRARY FEDERATION. n.d. ***DLF eResearch Network.*** Washington, DC: Digital Library Federation. [Online] available at https://www.diglib.org/groups/e-research-network/ (Accessed 3 September 2017).

DODGSON, M. 1996. Learning, trust and inter-firm technological linkages: some theoretical associations. In: ***Technological Collaboration: the dynamics of cooperation in industrial innovation.*** Edited by Rod Coombs, Albert Richards, Pier Paolo Saviotti, and Vivien Walsh. Cheltenham, UK; Brookfield, Vt.: Edward Elgar, p.54-75.

DONNELY, M. 2013. The DCC's institutional engagements: raising research data management capacity in UK higher education. ***Bulletin of the Association for Information Science and Technology***, August / September, 39(6): 37-40.

DONNELLY, M. 2015. Research data management and the H2020 Open Data Pilot. Presentation delivered at the ***FOSTER Event***, 22 October, University of Cyprus, Nicosia, Cyprus. [Online] available at https://www.fosteropenscience.eu/sites/default/files/pdf/1914.pdf (Accessed 3 January 2017).

Dropbox, 2012***.*** San Francisco, CA: Dropbox Inc. [Online] available at https://www.dropbox.com/ (Accessed 16 October 2012).

Drupal, n.d. Berchem, Belgium: Dries Buytaert BVBA. [Online] available at http://drupal.org/ (Accessed 6 October 2012).

DSpace, n.d. Winchester, MA: DuraSpace. [Online] available at http://www.dspace.org (Accessed 19 October 2012).

DSTC (DISTRIBUTED SYSTEMS TECHNOLOGY CENTRE). 2004. ***e-Research Middleware: The Missing Link in Australia's e-Research Agenda.*** Discussion whitepaper for submission to The Commonwealth of Australia, Department of Education, Science and Training, National Research Infrastructure Task Force. [Brisbane, QLD]: *Distributed Systems Technology Centre*, University of Queensland. [Online] available at: http://itee.uq.edu.au/~eresearch/papers/eResearchMiddleware.pdf (Accessed 9 February 2013).

DUNN, S. 2009. Dealing with the complexity deluge: VREs in the arts and humanities, ***Library Hi Tech***, 27(2): 205-216.

DURHAM UNIVERSITY. n.d. ***Why manage research data?*** Durham, UK: Research Office, Durham University. [Online] available at https://www.dur.ac.uk/research.office/research-outputs/research-data-management/why/ (Accessed 21 October 2014).

DUTTON, W.H. AND JEFFREYS, P.W. 2010. World wide research: an introduction. In: ***World wide research: reshaping the sciences and humanities.*** Edited by William H. Dutton and Paul W. Jeffreys. Cambridge, MA: MIT Press.

eCAT, 2011. Edinburgh, UK: Axiope. [Online] available at http://www.axiope.com/ (Accessed 3 October 2012).

EDINA, n.d. Edinburgh, UK: University of Edinburgh. [Online] available at https://edina.ac.uk/ (Accessed 24 September 2017).

Edinburgh University Data Library Research Data Management Handbook, v.1.0, August, 2011. [Edinburgh, UK: University of Edinburgh]. [Online] available at http://www.docs.is.ed.ac.uk/docs/data-library/EUDL_RDM_Handbook.pdf (Accessed 19 August 2014).

EndNote, 2012. New York, NY: Thomson Reuters. [Online] available at http://endnote.com/ (Accessed 15 October 2012).

ENDRES, A. AND ROMBACH, D. 2003. ***A handbook of software and systems engineering: empirical observations, laws and theories.*** Essex, UK: Pearson Educational.

EPRC. 2013. ***EPSRC Policy Framework on Research Data.*** Swindon, UK: Engineering and Physical Sciences Research Council. [Online] available at http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx (Accessed 22 March 2014).

eResearch Africa 2013 Conference, 2013***.*** Held 6-10 October, Southern Sun Newlands, Cape Town. [Online] available at http://eresearch.ac.za/ (Accessed 11 March 2014).

eResearch Africa 2017 Conference, 2017. Held 2-5 May, University of Cape Town, Cape Town. [Online] available at http://www.eresearch.ac.za/ (Accessed 16 September 2017).

E-Research: transformation in scholarly practice. Edited by Nicholas W. Jankowski. New York, NY: Routledge, 2009.

eRIC: e-Research Infrastructure and Communication, 2017. [Sl.: S.n.]. [Online] available at https://www.eric-project.org/ (Accessed 16 September 2017).

EsciDoc, 2012. [Sl: sn]. [Online] available at https://www.escidoc.org/ (Accessed 6 February 2013).

ESFRI. 2011. ***Strategy report on research infrastructures: Roadmap 2010.*** [Sl.]: European Union. [Online] available at http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf (Accessed 9 March 2014).

ESFRI. 2016. ***Strategy Report on research infrastructures: Roadmap 2016.*** [Sl.]: European Strategy Forum on Research Infrastructures. [Online] available at http://www.esfri.eu/roadmap-2016 (Accessed 13 December 2016).

Essential guide to API management and application integration, 2017. Newton, MA: TechTarget. [Online] available at http://searchmicroservices.techtarget.com/essentialguide/Essential-guide-to-API-management-and-application-integration (Accessed 29 August 2017).

Essex receives £5 million for new Big Data Network centre, 2013. ***Blogs Essex Daily***, web log post, 10 October 2013. [Online] available at http://blogs.essex.ac.uk/essexdaily/2013/10/10/big_data_network/ (Accessed 24 September 2017).

EUROPEAN COMMISSION. 2016a. ***European Open Science Cloud.*** [Sl.]: European Commission. [Online] available at http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud (Accessed 8 August 2016).

EUROPEAN COMMISSION. 2016b. ***The European Cloud Initiative.*** [Sl.]: European Commission. [Online] available at https://ec.europa.eu/digital-single-market/en/european-cloud-initiative (Accessed 9 May 2016).

EUROPEAN COMMISSION. 2016c. ***Guidelines on FAIR data management in Horizon 2020***. [Sl.]: European Commission. [Online] available at http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 19 September 2017).

EUROPEAN DATA PORTAL. 2016. ***What we do.*** [Sl.]: European Data Portal. [Online] available at https://www.europeandataportal.eu/en/what-we-do/our-activities (Accessed 19 September 2017).

EUROPEAN PARLIAMENT. 2016. ***Report on Towards a Digital Single Market Act***. (2015/2147(INI)) [Sl.]: European Parliament. [Online] available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2015-0371+0+DOC+XML+V0//EN (Accessed 19 September 2017).

EUDAT. n.d. ***What is EUDAT?*** [Sl.]: EUDAT. [Online] available at https://eudat.eu/what-eudat (Accessed 19 September 2017).

Evaluation toolbox, 2010. [Sl.: s.n.] [Online] available at http://evaluationtoolbox.net.au/ (Accessed 16 April 2015).

Evernote, 2012. Redwood City, CA: Evernote Corporation. [Online] available at http://evernote.com/ (Accessed 15 October 2012).

FALS-BORDA, O. 2000. Participatory (action) research in social theory: origins and challenges. In: ***Handbook of action research: participative inquiry and practice.*** Edited by Peter Reason and Hilary Bradbury. London: Sage, p.27-37.

FEARON, D., et al. 2013. Research data management services. ***SPEC Kit 334***, July 2013. Washington, D.C.: Association of Research Libraries. [Online] available at http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/ (Accessed 3 September 2017).

FedoraCommons, n.d. Winchester, MA: DuraSpace. [Online] available at http://fedora-commons.org/ (Accessed 9 August 2012).

FERNIHOUGH, S. 2011. ***E-Research: an implementation framework for South African organisations.*** MBA Research Report. Pretoria: University of South Africa. [Online] available at http://hdl.handle.net/10500/4474 (Accessed 2 June 2012).

FILETTI, M. & GNAUCK, A. 2011. A concept of a Virtual Research Environment for long-term ecological projects with free and open source software. In: ***Environmental Software Systems: frameworks of eEnvironment.*** (IFIP Advances in Information and

Communication Technology, 359), Proceedings of the 9th IFIP WG 5.11 International Symposium, ISESS 2011, 27-29 June, Brno, Czech Republic. Edited by Jiří Hřebíček, Gerald Schimak, and Ralf Denzer. Berlin, Heidelberg: Springer, p. 235-244.

Flickr, 2012 Sunnyvale, CA: Yahoo Inc. [Online] available at http://www.flickr.com (Accessed 20 October 2012).

FLORES, J.R., BRODEUR, J.J., DANIELS, M.G., NICHOLLS, N. & TURNATOR, E. 2015. Libraries and the research data management landscape. In: ***The process of discovery: the CLIR Postdoctoral Fellowship Program and the future of the academy.*** Edited by John C. Maclachlan, Elizabeth A. Waraksa, and Christa Williford. Washington, DC: Council on Library and Information Resources. [Online] available at https://www.clir.org/pubs/reports/pub167/RDM.pdf (Accessed 25 September 2017).

FLOYD, C. 1984. A systematic look at prototyping. In: ***Approaches to prototyping.*** Edited by R. Budde, K. Kuhlenkamp, L. Mathiassen, and H. Zűllighoven. Berlin: Springer-Verlag, p.1-18.

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

FOSTER, E.D. & DEARDORFF, A. 2017. Open Science Framework (OSF). ***Journal of Medical Library Association***, April, 105(2): 203-206. [Online] available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370619/ (Accessed 12 February 2018).

FOSTER, I., KESSELMAN, C. & TUECKE, S. 2001. The anatomy of the grid: enabling scalable virtual organizations. The International Journal of Supercomputer Applications, 15(3): 200-202.

Fourth paradigm: data-intensive scientific discovery. Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Redmond, WA.: Microsoft Research, c2009.

FRASER, M. 2005. Virtual research environments: overview and activity. July, ***Ariadne***, Issue 44. [Online] available at http://www.ariadne.ac.uk/issue44/fraser/ (Accessed 30 October 2011).

FRICKE, M. 2009. The knowledge pyramid: a critique of the DIKW hierarchy. ***Journal of Information Science***, 23 October, 35(2): 131-142.

FRIEDLANDER, A. & ALDER, P. 2006. ***To stand the test of time: long-term stewardship of digital data sets in science and engineering: a report to the National Science Foundation from the ARL workshop on new collaborative relationships: the role of academic libraries in the digital data universe.*** Washington, DC: Association of Research Libraries. [Online] available at http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf (Accessed 3 September 2017).

FRIENDLY, M. 2009. ***Milestones in the history of thematic cartography, statistical graphics, and data visualization.*** [Sl.: s.n.] [Online] available at http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf (Accessed 19 September 2014).

FROSINI, L. 2016. ***GCube reference archictecture.*** [Online] available at https://gcube.wiki.gcube-system.org/gcube/GCube_Reference_Architecture (Accessed 12 February 2018).

FRY, J. & SCHROEDER, R. 2009. Towards a sociology of e-research: shaping practice and advancing knowledge. In: ***E-Research: transformation in scholarly practice.*** Edited by Nicholas Jankowski. New York and London: Routledge.

FULLARD, A. 2016. **RDM activities at UWC.** Presented [online] at the NeDICC meeting, 18 February, CSIR, Pretoria.

GERGEN, K.J. 1999. ***An invitation to social construction.*** London: Sage.

GERRING, J. 2007. ***Case study research: principles and practices.*** New York, NY: Cambridge University Press.

GILL, P., STEWART, K., TREASURE, E. & CHADWICK, B. 2008. Methods of data collection in qualitative research: interviews and focus groups. ***British Dental Journal*,** 22 March, 204(6): 291-295. [Online] available at http://www.nature.com/bdj/journal/v204/n6/full/bdj.2008.192.html (Accessed 16 September 2014).

GILLIS, A. AND JACKSON, W. 2002. ***Research for nurses: methods and interpretation.*** Philadelphia, PA: F.A. Davis Company.

GLAVES, H. 2016. ***VRE-IG charter.*** [Sl.]: Research Data Alliance. [Online] available at https://www.rd-alliance.org/system/files/documents/BoF%20VREs%20Charter.pptx (Accessed 21 September 2017).

GOEBBELS, G. & LALIOTI, V. 2001. Co-presence and co-working in distributed collaborative environments. In: Afrigraph 01': proceedings of the 1st International conference on Computer graphics, virtual reality and visualisation, Camps Bay, Cape Town, south South Africa, 5-7 November 2001, p. 109-114.

Google, n.d. Mountain View, CA: Google Inc. [Online] available at http://www.google.com (Accessed 20 October 2012).

Google Apps for Education. Mountain View, CA: Google Inc, n.d. [Online] available at http://www.google.com/enterprise/apps/education/ (Accessed 24 March 2013).

Google Drive. Mountain View, CA: Google Inc., 2013. [Online] available at http://drive.google.com (Accessed 22 March 2013).

Google+ Hangouts. Mountain View, CA: Google Inc., 2012. [Online] available at https://tools.google.com/dlpage/hangoutplugin (Accessed 16 October 2012).

Google Talk. Mountain View, CA: Google Inc., 2011. [Online] available at http://www.google.com/talk/about.html (Accessed 15 October 2012).

GRADY, J. 2008. Virtual research at the crossroads. ***Forum Qualitative Sozialforschung / Forum: Qualitative Social Research***, 9(3), Art. 38. [Online] available at http://nbn-resolving.de/urn:nbn:de:0114-fqs0803384 (Accessed 9 May 2015).

GRAZIANO, A.M. & RAULIN, M.L. 2000. ***Research methods: a process of inquiry*****.** 4th ed. Boston: Allyn and Bacon, Boston.

GUIDA, G., LAMPERTI, G. & ZANELLA, M. 1999. ***Software prototyping in data and knowledge engineering.*** Dordrecht: Springer Science & Business Media.

HABERMAS, J. 1997. ***Between facts and norms: contributions to a discourse theory of law and democracy.*** Cambridge, UK: Polity Press.

HAINES, R. 2012. Data Governance is strategic, data stewardship is tactical. ***EMC-In Focus***, 3 May 2012. [Online] available at https://infocus.emc.com/rachel_haines/data-governance-is-strategic-data-stewardship-is-tactical/ (Accessed 18 August 2014).

HALBERT, M. 2013. Prospects for research data management. In: ***Research data management: principles, practices and prospects.*** Washington, D.C.: DataRes, Council on Library and Information Resources, p.1-15. [Online] available at https://www.clir.org/pubs/reports/pub160/pub160.pdf (Accessed 24 September 2017).

HALFPENNY, P., PROCTOR, R., LIN, Y. & VOSS, A. 2009. Developing the UK-based e-Social Science Research Program. In: ***E-Research: transformation in scholarly practice.*** Edited by Nicholas W. Jankowski. New York, NY: Routledge, p. 74-90.

HAMMOND, M. 2017. ***Next Generation Research Environments: discovery phase report JSC1701D001-1.0.*** [Sl.: s.n.] [Online] available at http://repository.jisc.ac.uk/6669/1/JSC1701D001-1.0_NGRE_Discovery_report.pdf (Accessed 21 September 2017).

Handbook of action research: participative inquiry and practice. Edited by Peter Reason and Hilary Bradbury. London: Sage.

HARA, N. & ROSENBAUM, H. 2008. Revising the conceptualization of Computerization Movements. ***The Information Society***, 24(4): 229-245.

HARVEY, R. 2010. ***Digital curation: a how-to-do-it manual.*** London: Facet Publishing.

HASTAC: Humanities, Arts, Science and Technology Advanced Collaboratory. [Online] available at http://hastac.org (Accessed 5 February 2012).

THE HATCHER GROUP. 2008. Glossary of new media and online outreach terminology. ***The Feed***, web log post 8 May 2008. [Online] available at https://thehatchergroup.wordpress.com/2008/05/08/glossary-of-new-media-terminology/ (Accessed 12 February 2018).

HAVENGA, H.M. 2008. ***An investigation of students' knowledge, skills and strategies during problem solving in object-oriented programming.*** Doctoral Thesis, University of South Africa. Pretoria: UNISA.

HAYES, H. 2005. Digital repositories: helping universities and colleges. ***JISC Briefing Paper: Higher Education Sector***. Edinburgh, UK: JISC. [Online] available at https://web.archive.org/web/20140718011638/http://www.jisc.ac.uk/uploaded_documents/HE_repositories_briefing_paper_2005.pdf (Accessed 12 February 2018).

HENDRIKSE, S. 2017. ***Stellenbosch University: Library and Information Service: Manager: Research Data Services.*** E-mail received on 10 April 2017.

HENTY, M., WEAVER, B., BRADBURY, S. & PORTER, S. 2008. ***Investigating data management practices in Australian universities.*** [Sl.]: Australian Partnership for Sustainable Repositories (APSR). [Online] avaialable at http://apsr.anu.edu.au/orca/investigating_data_management.pdf (Accessed 9 September 2017).

HEY, A.J.G. & TREFETHEN, A.E. 2003. The data deluge: an e-science perspective. In: Berman, F. et al., ***Grid Computing - Making the Global Infrastructure a Reality,*** *edited by Fran Berman, Geoffrey Fox and Tony Hey.* Chichester, UK: Wiley and Sons, p. 809-824.

HEY, T. & HEY, J. 2006. E-Science and its implications for the library community. ***Library Hi Tech,*** 24(4): 515-528.

HEY, T. & TREFETHEN, A. 2008. E-Science, cyberinfrastructure, and scholarly communication. In: ***Scientific collaboration on the Internet.*** Edited by Gary M. Olson, Ann Zimmerman, and Nathan Bos. Cambridge, MA: The MIT Press, p.15-31.

HIGH LEVEL EXPERT GROUP ON SCIENTIFIC DATA. 2010. ***Riding the wave: how Europe can gain from the rising tide of scientific data: final report by the High Level Expert Group on Scientific Data: a submission to the European Commission.*** [Sl.]: European Union. [Online] available at http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707 (Accessed 3 January 2017).

HINE, C. 2006. Computerization Movements and scientific disciplines: the reflexive potential of new technologies. In: ***New infrastructures for knowledge production: understanding e-science.*** Edited by C. Hine. London: Information Science Publishing, p.26-47.

HINE, C. 2008. ***Systematics as cyberscience: computers, change, and continuity in science.*** Cambridge, MA: MIT Press.

HORTON, L., VAN DEN EYNDEN, V., CORTI, L. & BISHOP, L. 2011. ***Data Management recommendations for research centres and programmes.*** Colcherster, UK: UK Data Archive. [Online] available at http://www.data-archive.ac.uk/media/257765/ukda_datamanagementrecommendations_centresprogrammes.pdf (Accessed 25 September 2017).

HUADONG, G. 2014*.* Scientific Big Data for knowledge discovery. Presented at the ***CODATA Workshop on Big Data for International Scientific Programmes***, 8-9 June, Beijing, China.

HUBZERO. 2017a. ***About us.*** West Lafayette, IN: Purdue University, HUBzero. [Online] available at https://hubzero.org/about (Accessed 22 September 2017)

HUBZERO. 2017b. ***Documentation.*** West Lafayette, IN: Purdue University, HUBzero. [Online] available at https://help.hubzero.org/documentation/current/user

HUGHS, B. & COTTERELL, M. 2002. ***Software project management.*** 3rd ed. Berkshire, UK: McGraw-Hill.

HUMPHREY, C. 2006. ***E-Science and the lifecycle of research.*** [Online] available at http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc (Accessed 23 August 2010).

HWANG, J., CAPELLAN, E.S. & CONSULTA, R.R. 2008. e-Science Models and the Research Life Cycle, how will it affect the Philippine Community? In: ***Proceedings of the 2008 International Conference on Grid Computing and Applications, GCA 2008, Las Vegas, Nevada, USA, July 14-17, 2008.*** Edited by Hamid R. Arabnia. [Las Vegas, Nevada]: CSREA Press, p. 61-76.

HWANG, K., DONGARRA, J. & FOX, G.C. 2013. ***Distributed and cloud computing; from parallel processing to the Internet of Things.*** Waltham, MA: Morgan Kaufmann

IACONO, S. & KLING, R. 2001. Computerization Movements: the rise of the Internet and distant forms of work. In: ***Information Technology and Organizational Transformation: History, Rhetoric and Practice,*** Edited by J.A. Yates and J. van Maanen. Thousand Oaks, CA: Sage Publications.

ICSU. n.d. ***World Data System***. Tokyo, Japan: ICSU, World Data System. [Online] available at https://www.icsu-wds.org/organization (Accessed 30 July 2014).

IDIA: Inter-University Institute for Data Intensive Astronomy, 2017. Cape Town: IDIA (Inter-University Institute for Data Intensive Astronomy). [Online] available at http://idia.ac.za/ (Accessed 21 January 2017).

IFDO: International Federation of Data Organisations for Social Science, n.d. [Sl.]: IFDO: International Federation of Data Organisations for Social Science. [Online] available at http://ifdo.org/wordpress/ (Accessed 4 January 2017).

INDIANA UNIVERSITY. 2012. ***On XSEDE, what is a Science Gateway?*** [Online] available at http://kb.iu.edu/data/ascm.html (Accessed 4 February 2012).

INFORMATICA. 2018. ***What is data validation?*** Redwood City, CA: Informatica. [Online] available at ***IFDO: International Federation of Data Organisations for Social Science***, n.d. [Sl.]: IFDO: International Federation of Data Organisations for Social Science. [Online] available at http://ifdo.org/wordpress/ (Accessed 4 January 2017).

INSTITUTE OF EDUCATION SCIENCES. n.d. ***Data sharing and implementation guide****.* Washington, DC: Institute of Education Sciences, U.S. Department of Education. [Online] available at http://ies.ed.gov/funding/datasharing_Implementation.asp (Accessed 19 September 2014).

INTERACTION DESIGN FOUNDATION. 2017. ***Design iteration brings powerful results: so do it again designer!*** Aarhus, Denmark: Interaction Design Foundation. [Online] available at https://www.interaction-design.org/literature/article/design-iteration-brings-powerful-results-so-do-it-again-designer (Accessed 24 September 2017).

Interview with F. van Till and M. Dovey, JISC, 1 June 2010, HEFC Building, London, UK.

IPUMS Terra, 2016. [Sl.]: Minnesota Population Center. [Online] available at https://www.terrapop.org/ (Accessed 2 September 2017).

ISI Web of Knowledge, 2013. New York, NY: Thomson Reuters.

JAMISON, S. AND BORTLIK, C. & HANLEY, S. 2013. Planning your SharePoint 2013 solution strategy, Informit, 9 October 2013. [Online] available at http://www.informit.com/articles/article.aspx?p=2130299&seqNum=2 (Accessed 12 February 2018).

JANKOWSKI, N. & CALDAS, A. 2004. E-Science: principles, projects and possibilities for communication and Internet studies. Paper presented at ***Etmaal van de Communicatiewetenschap (Day of Communication Science)***, 18-19 November, University of Twente, the Netherlands.

JANKOWSKI, N.W. 2007. Exploring e-science: an introduction. ***Journal of Computer-Mediated Communication***, 12(2): 549-562.

JANKOWSKI, N.W. 2009. The contours and challenges of e-research. In: ***E-Research: transformation in scholarly practice.*** Edited by Nicholas W. Jankowski. New York, NY: Routledge, p. 3-34.

Jchem, n.d. Budapest, Hungary: ChemAxon Ltd. [Online] available at http://www.chemaxon.com/jchem/intro/index.html (Accessed 22 March 2013).

JEFFREYS, P.W. 2010. The developing conception of e-research. In: ***World wide research: reshaping the humanities.*** Edited by William H. Dutton & Paul W. Jeffreys. Cambridge, MA: The MIT Press, p. 51-66.

Jim Gray on eScience: a transformed scientific method. In: ***Fourth paradigm: data-intensive scientific discovery.*** Edited by Tony Hey, Stewart Tansley, & Kristin Tolle. Redmond, WA: Microsoft Research, c2009, p. xvii-xxxi.

JISC. n.d. ***Research at Risk.*** UK: Joint Information Systems Committee (JISC). [Online] available at http://www.jisc.ac.uk/rd/projects/research-at-risk (Accessed 1 September 2017).

JISC. 2006. ***JISC Virtual Research Environments Programme: Phase 2 Roadmap***. London, UK: Joint Information Systems Committee (JISC). [Online] available at http://www.jisc.ac.uk/publications/programmerelated/2006/pub_vreroadmap.aspx (Accessed 17 February 2012).

JISC. 2012. ***JISC Virtual Research Environments Programme***. London, UK: Joint Information Systems Committee (JISC). [Online] available at http://www.jisc.ac.uk/programme_vre.html (Accessed 17 February 2012).

JISC. 2014a. ***Managing research data.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at https://www.jisc.ac.uk/rd/projects/managing-research-data (Accessed 30 August 2017).

JISC. 2014b. ***Virtual Research Environment Programme.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://webarchive.nationalarchives.gov.uk/20140702163345/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx (Accessed 21 September 2017).

JISC. 2014c. ***Digital Infrastructure: Research Programme.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://webarchive.nationalarchives.gov.uk/20140702162920/http://www.jisc.ac.uk/whatwedo/programmes/di_research.aspx (Accessed 21 September 2017).

JISC. 2014d. ***Research Programme: Research Tools.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://webarchive.nationalarchives.gov.uk/20140702163250/http://www.jisc.ac.uk/whatwedo/programmes/di_research/researchtools.aspx (Accessed 21 September 2017).

JISC. 2014e. ***Research Programme: Research Support.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://webarchive.nationalarchives.gov.uk/20140702163249/http://www.jisc.ac.uk/whatwedo/programmes/di_research/researchsupport.aspx (Accessed 21 September 2017).

JISC. 2014f. ***JISC Virtual Research Environments programme (Phase 1)***. London, UK: Joint Information Systems Committee (JISC). [Online] available at http://www.webarchive.org.uk/wayback/archive/20140614021919/http://www.jisc.ac.uk/whatwedo/programmes/vre1.aspx (Accessed 5 February 2015).

JISC. 2016a. ***About us.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://www.jisc.ac.uk/about (Accessed 1 December 2016).

JISC. 2016b. ***Implementing a virtual research environment (VRE): understanding the tools and technologies needed by researchers.*** London, UK: Joint Information Systems Committee (JISC). [Online] available at http://www.jisc.ac.uk/full-guide/implementing-a-virtual-research-environment-vre (Accessed 12 February 2018).

JOHNSSON, M. & ÅHLFELDT, J. 2015. ***Research libraries and research data management within the humanities and social sciences: project report.*** Lund, Sweden: Lund University Library, Department of Research and Study Services. [Online] available at http://lup.lub.lu.se/record/5050462 (Accessed 2 January 2017).

KAASE, M. 2001. Databases, core: political science and political behavior. In: ***International Encyclopedia of the Social and Behavioral Sciences, Vol. 5.*** Edited by N.J. Smelser & P.B. Baltes. Amsterdam: Elsevier, p. 3251-3255.

KAHN, S. 2004. The future of computational science. **Scientific Computing World,** May / June. [Online] available at http://www.scientific-computing.com/features/feature.php?feature_id=86 (Accessed 21 January 2012).

KALLENBORN, R. 2013. eRIC: E-Research - Infrastructure and Communication: a New Leaf for Library Services. Presented at ***UFSC / IATUL Workshop, 'Challenges of Networking Library Services'***, 30 September-1 October, Florianopolis, Santa Catarina, Brazil. [Online] available at http://workshopbuiatul.ufsc.br/files/2013/10/IATUL-Florianopolis-2013-a-New-Leaf.pdf (Accessed 30 April 2014).

KARLSEN, J.I. 1991. Action research as method: reflections from a program for developing methods and competence. In: ***Participatory action research.*** Edited by William Foote Whyte. (Sage focus edition; vol. 123). Newbury Park, CA: Sage Publications, p.143-158.

KASHYAP, V. 2010. What is an API and what are they good for? ***MakeUseOf***, 23 August 2010. [Online] available at http://www.makeuseof.com/tag/api-good-technology-explained/. (Accessed 10 October 2012).

KENNAN, M.A. & MARKAUSKAITE, L. 2015. Research data management practices: a snapshot in time. ***International Journal of Digital Curation***, 10(2): 69-95. [Online] available at http://www.ijdc.net/index.php/ijdc/article/view/10.2.69 (Accessed 12 September 2017).

The Kepler project, n.d. [S.l.]: Kepler. [Online] available at https://kepler-project.org/ (Accessed 9 August 2012).

KERAMINIYAGE, K., AMARATUNGA, D. & HAIGH, R. 2009a. Achieving success in collaborative research: the role of Virtual Research Environments. ***ITcon,*** 14(Special Issue): 59-69. [Online] available at http://www.itcon.org/data/works/att/2009_07. content.05004.pdf (Accessed 26 February 2012).

KERAMINIYAGE, K., AMARATUNGA, D. & HAIGH, R. 2009b. A human-computer interaction principles based framework to assess the user perception of web based virtual research environments. ***International Journal of Strategic Property Management***, 13: 129-142.

KINGSLEY, D. 2016. Consider yourself disrupted: notes from RLUK2016. ***Unlocking Research,*** web log post, 14 March 2016. [Cambridge, UK]: University of Cambridge, Office of Scholarly Communication. [Online] available at https://unlockingresearch. blog.lib.cam.ac.uk/?p=601 (Accessed 12 September 2017).

KLAPWIJK, W. 2014. ***Ontwikkelinge rondom RDM by Stellenbosch Universiteit.*** E-mail received on 19 May 2014.

KLEIN, H.K. & MYERS. 1999. A set of principles for conducting and evaluating interpretive field studies in information systems. ***MIS Quarterly***, March, 23(1): 67-94.

KLING, R. & IACONO, S. 1988. The Mobilization of Support for Computerization: the role of Computerization Movements. ***Social Problems**,* 35**:** 226-243.

KLYNE, G. 2006. Sakai VRE Demonstrator project user requirements. ***OSS Watch Wiki***, wiki article, 22 February 2006. [Online] available at http://wiki.oss-watch.ac.uk/SakaiVre/UserRequirements (Accessed 2 September 2012).

KOCK, N., DAVISON, R., OCKER, R. & WAZLAWICK, R. 2001. E-collaboration: a look at past research and future challenges. ***Journal of Systems and Information. Technology***, 5(1): 1–9.

KRAUT, R., EGIDO, C. & GALEGHER, J. 1988. Patterns of contact and communication in scientific research collaboration. In: ***CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work.*** New York, NY: ACM, p. 1-12.

KUMAR, S. 2009. Interoperability protocols and standards in LIS*.* Presented at the ***National Workshop on Library 2.0: A Global Information Hub***, Physical Research Laboratory Ahmedabad, India. [Online] available at http://www.slideshare.net/alibnetweb/interoperability-protocols-and-standards-in-lis (Accessed 6 October 2012).

KVALHEIM, V. & KVAMME, T. 2014. ***Policies for sharing research data in social sciences and humanities: a survey about research funders' data policies.*** [Sl.]: International Federation of Data Organizations for Social Science. [Online] available at http://www.ada.edu.au/documents/ifdo-report-on-policies-for-data-sharing (Accessed 4 January 2017).

LANTZ, K.E. n.d. ***The prototyping methodology.*** Englewood Cliffs, NJ: Prentice-Hall.

LAVE, J. & WENGER, E. 1991. ***Situated Learning: legitimate peripheral participation.*** Cambridge, UK: Cambridge University Press.

LAWSON, I. & BUTSON, R. 2007. ***eResearch at Otago: a report of the University of Otago eResearch Steering Group***. Dunedin, New Zealand: University of Otago. [Online] available at https://web.archive.org/web/20100526002658/http://eresearch.wiki.otago.ac.nz/images/0/0d/EResearch%40Otago_Report_07.pdf (Accessed 28 February 2015).

LEE, A.S. 1989. A scientific methodology for MIS case studies. ***MIS Quarterly***, March, 13(1): 33-50.

LEONARDO, C., CASTELLI, D. & PAGANO, P. 2009. On-demand virtual research environments and the changing roles of librarians. ***Library Hi Tech***, 27(2): 239-251.

LERU RESEARCH DATA WORKING GROUP. 2013. ***LERU Roadmap for Research Data.*** (LERU Advice Papers 14). LERU: Leuven, Belgium. [Online] available at http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf (Accessed 2 January 2017).

LESHEM, S. & TRAFFORD, V. 2007. Overlooking the conceptual framework. ***Innovations in Education and Teaching International,*** February, 44(1): 93-105.

LEVINE, J.R., YOUNG, M.L. & BAROUDI, C. 2005. ***The Internet for dummies.*** 10th ed. Hoboken, NJ: Wiley.

LIASA HELIG and WC HELIG WORKSHOP ORGANIZING COMMITTEE, DAVIDSON, J. & JONES, S. 2014. ***LIASA Report on Developing Research Data Management Services Workshop***, held 27 March 2014, CPUT Library, Bellville Campus, Cape Town, South Africa.

LIASA WCHELIG/HELIG/DCC Workshop on Developing Research Data Management Services, 27 March 2014, Cape Town. [Unpublished].

LICHTER, H., SCHNEIDER-HUFSCHMIDT, M. & ZÜLLIGHOVEN, H. 1994. Prototyping in industrial software projects: bridging the gap between theory and practice. ***IEEE Transactions on Software Engineering***, November, 20(11): 825-832.

Liferay, 2012. [Los Angeles, CA]: Liferay Inc. [Online] available at http://www.liferay.com/ (Accessed 9 August 2012).

LINCOLN, Y.S. 2001. Engaging sympathies: relationships between action research and social constructivism. In: ***Handbook of action research: participative inquiry and practice.*** Edited by Peter Reason and Hilary Bradbury. London: Sage, p.124-132.

LinkedIn, 2013. Mountain View, CA: LinkedIn Corporation. [Online] available at http://www.linkedin.com (Accessed 22 March 2013).

LORD, P. & MACDONALD, A. 2003. ***Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision.*** Prepared for: The JISC Committee for the Support of Research (JCSR) (e-Science Curation Report). Twickenham, UK: The Digital Archiving Consultancy Limited. [Online] available at http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf (Accessed 18 August 2014).

LÖTTER, L. 2014a. ***Databasis / sagteware wat julle gebruik vir julle data***. E-mail received 13 March 2014.

LÖTTER, L. 2014b. Reflections on the RDM position in South Africa. Presented at ***Developing Research Data Management Services, LIASA WCHELIG/HELIG/DCC Workshop,*** 27 March, CPUT Library, Bellville Campus, Cape Town, South Africa.

LÖTTER, L. & VAN ZYL, C. 2015. A reflection on a data curation journey. ***Journal of Empirical Research on Human Research Ethics***, July, 10(3): 338-343.

LPDS. 2013. ***Grid and cloud user support environment.*** Budapest: Laboratory of Parallel and Distributed Systems (LPDS), Institute for Computer Science and Control, Hungarian Academy of Science. [Online] available at http://guse.hu/about/home (Accessed 12 February 2015).

LYNCH, C. 2008. The institutional challenges of Cyberinfrastructure and E-Research. ***Educause Review***, Nov / Dec, 43(6): 74-88.

MA, L. 2012. Meanings of information: the assumptions and research consequences of three foundational LIS theories. ***Journal of the American Society for Information Science and Technology***, April, 63(4): 716-723.

MACANDA, M., RAMMUTLOA, M. & BEZUIDENHOUT, R. 2014. ***Research Data Management at UNISA.*** Paper presented at the ***Annual LIS Research Symposium***, 24-25 July, UNISA Science Campus, Florida, South Africa. [Online] available at http://hdl.handle.net/10500/13907 (Accessed 5 January 2017).

MACGREGOR, K. 2015. Seeking 200 data scientists for huge radio telescope. ***University World News***, Global Edition, 6 September 2015, Issue 380. [Online] available at http://www.universityworldnews.com/article.php?story=20150906085207602 (Accessed 16 September 2017).

MACHLUP, F. 1984. Semantic quirks in studies of information. In: ***The study of information: interdisciplinary messages.*** Edited by F. Machlup and U. Mansfield. New Yorl, NY: Wiley, p. 641-671.

MACK, N., WOODSONG, C., MACQUEEN, K.M., GUEST, G. & NAMEY, E. 2005. ***Qualitative research methods: a data collector's field guide.*** Research Triangle Park, NC: Family Health International. [Online] available at http://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector%27s%20Field%20Guide.pdf (Accessed 17 September 2014).

MAKOLA, D., SIM, Y.W., WANG, C., GILBERT, L., GRANGE, S. & WILLS, G. 2006. A Service-Oriented Architecture for a Collaborative Orthopaedic Research Environment. In: ***8th Annual Conference on WWW Applications,*** 6 to 8 September 2006, Bloemfontein, South Africa. [Online] available at http://eprints.ecs.soton.ac.uk/12898/1/WWW2006Core.pdf (Accessed 28 January 2012).

Managing research data. Edited by Graham Pryor. London: Facet Publishing, 2012.
MARSHALL, B. 2002. Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative**. *Computers in Libraries***, September, 22(8): 24-29. [Online] available at https://librarytechnology.org/document/9944 (Accessed 29 August 2017).

MARSHALL, C. & ROSSMAN, G.B. 2006. ***Designing qualitative research.*** 4th ed. Thousand Oaks, CA: Sage.

MARTIN, A. 2014. What is the difference between a framework and a platform? ***Stackoverflow***, web log post, 30 July 2014. [Online] available at https://stackoverflow.com/questions/25028243/what-is-the-difference-between-a-framework-and-a-platform (Accessed 12 February 2018).

MARTIN, E. & BALLARD, G. 2010. ***Data management best practices and standards for Biodiversity data applicable to Bird Monitoring Data.*** U.S. North American Bird Conservation Initiative Monitoring Subcommittee. [Online] available at http://www.nabci-us.org/ (Accessed 24 September 2014).

MARTINEZ-URIBE, L. & MACDONALD. S. 2009. User engagement in research data ciration. In: ***Research and Advanced Technology for Digital Libraries: 13th European Conference, ECDL 2009, Corfu, Greece, 27 September-2 October 2009: proceedings.*** Edited by Agosti, M., Borbinha, J.L., Kapidakis, S., Papatheodorou, C. & Tsakonas, G. Berlin, Heidelberg: Springer, p. 309-314.

MATLATSE, R., PIENAAR, H. & VAN DEVENTER, M. 2017. ***Mobilising a nation: RDM training and education in South Africa.*** Presented by Heila Pienaar at the ***12th International Digital Curation Conference***, 20-23 February, Edinburgh, UK. [Onine] available at http://www.dcc.ac.uk/sites/default/files/documents/IDCC17~/presentations/IDCC17Training%20in%20SA%20rev%20MvDHeila.pdf (Accessed 17 September 2017).

MAXWELL, J. A. 1996. ***Qualitative research design: an interactive approach.*** Thousand Oaks, CA: Sage.

MAYKUT, P. & MOREHOUSE, R. 1994. ***Beginning qualitative research: a philosophic and practical guide.*** London: Farmer Press.

MCDONALD, C. 2005. Research as an information systems domain. In: ***Information systems foundations: constructing and criticising.*** Edited by Denis Hart and Shirley Gregor, 2005. Canberra: Australian National University E Press, p.145-151. [Online] available at: http://epress.anu.edu.au/info_systems/mobile_devices/index.html (Accessed 21 January 2012).

MCLENNAN, M. & KENNELL, R. 2010. HUBzero: a platform for dissemination and collaboration in computational science and engineering. ***Computing in Science and Engineering***, March-April, 12(2): 48-53.

MCLENNAN, M. & KLINE, G. 2011. HUBzero paving the way for the Third Pillar of Science. ***HPC Wire***, 28 February. [Online] available at https://www.hpcwire.com/2011/02/28/hubzero_paving_the_way_for_the_third_pillar_of_science/ (Accessed 24 September 2017).

MCTAGGART, R. 1991. Principles for participatory action research. ***Adult Education Quarterly***, Spring, 41(3): 168-187.

MIAS, E. 2016. ***Digital Library Services at UCT Libraries.*** Presented online at the ***NeDiCC meeting***, 18 February, CSIR, Pretoria.

MICHELINI, A. & LECARPENTIER, D. 2011. ***The EUDAT project: towards a European collaborative data infrastructure.*** Presented at the VERCE Kick-off, 3 October, Paris, France. [Online] available at http://www.verce.eu/Kickoff/Session1/VERCE-EUDAT.pdf (Accessed 25 September 2017).

Mendeley. London, UK: Mendeley Ltd, 2012. [Online] available at http://www.mendeley.com/ (Accessed 15 October 2012).

MICROSOFT TECHNET. 2015. ***Numeric data.*** [Sl.]: Microsoft. [Online] available at https://technet.microsoft.com/en-us/library/aa933110%28v=sql.80%29.aspx (Accessed 9 May 2015).

MOHOLOLA, E. 2016. ***UCT led consortium to build first regional data node of national cyberinfrastructure.*** UCT Press release, 17 August 2016. Cape Town: University of Cape Town. [Online] available at http://www.uct.ac.za/usr/press/2016/2016-08-17_Release_DST_Aug2016_EM.pdf (Accessed 16 September 2017).

Moodle. Perth, Australia: Moodle Pty Ltd, 2009. [Online] available at http://moodle.org/ (Accessed 28 September 2012).

MORSE, R., NADKARNI, P., SCHOENFELD, D.A. & FINKELSTEIN, D.M. 2011. Web-browser encryption of personal health information. ***BMC Medical Informatics and Decision Making***, 10 November, 11: 1-9. [Online] available at https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-11-70 (Accessed 30 March 2017).

MOSCOVE, S.A. 2001. Prototyping: an alternative approach to systems development work. ***Review of Business Information Systems***, 5(3): 65-72.

MOSSINK, W., BIJSTERBOSCH, M. & NORTIER, J. 2013. ***European landscape study of research data management: for SIM4RDM-Support Infrastructure Models for Research Data Management.*** Utrecht, Netherlands: SURF. [Online] available at http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20final%2025.01.12.pdf (Accessed 19 March 2014).

MS Project, n.d. Redmond, WA: Microsoft Corporation. [Online] available at http://office.microsoft.com/en-us/project/ (Accessed 23 March 2013).

MS Word, n.d. Redmons, WA: Microsoft Corporation. [Online] available at http://office.microsoft.com/en-za/word/ (Accessed 22 March 2013).

MOUTON, J. 2001. ***How to succeed in your master's and doctoral studies: a South African guide and resource book.*** Pretoria: Van Schaik.

MÜLLER, H. 2009. Database Archiving. In: ***DCC Briefing Papers: Introduction to Curation***. Edinburgh: Digital Curation Centre. [Online] available at: http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/database-archiving (Accessed 19 August 2014).

MÜNCH, V. 2011. The cradle of e-research: worldwide interconnected working environments**. *Online***, March-April, 35(2): 30-33.

MURRAY, J.D. & VANRYPER, W. 1996. Types of data. In: Murray, J.D. and VanRyper, W., ***Encyclopedia of Graphics File Formats***. 2nd. Ed. (O'Reiley Series). Sebastopol, CA: O'Reilly and Associates. [Online] available at http://www.fileformat.info/mirror/egff/ch10_03.htm (Accessed 10 May 2015).

MURRAY-RUST, P., NEYLON, C., POLLOCK, R. & WILBANKS, J. 2010. ***Panton principles: principles for open data in science.*** [Sl.: s.n.]. [Online] available at http://pantonprinciples.org/ (Accessed 3 December 2016).

myExperiment, 2010. Manchester, UK: University of Manchester; Southampton, UK: University of Southampton. [Online] available at http://www.myexperiment.org/ (Accessed 9 August 2012).

MYHILL, M., SHOEBRIDGE, M. & SNOOK, L. 2009. Virtual research environments: a Web 2.0 cookbook? ***Library Hi Tech***, 27(2): 228-238.

nanoHUB, 2017a. [Sl.: s.n.]. [Online] available at https://nanohub.org/. (Accessed 25 September 2017).

NANOHUB. 2017b. ***Add Rappture to your software development: learning module.*** [Sl.]: NanoHUB. [Online] available at https://nanohub.org/resources/240 (Accessed 22 September 2017).

The National Data Service: a vision for accelerating discovery through data sharing, 2014. (Version 3.0.1.) [Sl.: s.n.]. [Online] available at http://www.nationaldataservice.org/NDS-Summary.pdf (Accessed on 21 July 2014).

National Integrated Cyber-Infrastructure System, 2017. Pretoria: CSIR. [Online] available at https://www.csir.co.za/national-integrated-cyber-infrastructure-system (Accessed 4 February 2017).

NATIONAL RESEARCH COUNCIL. 1997. ***Bits of power: issues in global access to scientific power.*** Washington DC: National Academies Press.

NATIONAL RESEARCH COUNCIL. 2010***. Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata.*** Edited by Robert M. Hauser, Maxine Weinstein, Robert Pool, and Barney Cohen. Washington, DC: The National Academies Press.

NATIONAL SCIENCE FOUNDATION (NSF). 2016. ***NSF commits $35 million to improve scientific software.*** Arlington, VA: National Science Foundation (NSF). [Online] available at https://www.nsf.gov/news/news_summ.jsp?cntn_id=189347 (Accessed 23 September 2017).

NATIONAL STANDARDS POLICY ADVISORY COMMITTEE. 1978. ***National policy on standards for the United States and a recommended implementation plan***. Washington, DC: National Standards Policy Advisory Committee.

NeDICC, n.d. [Online] available at http://www.nedicc.ac.za (Accessed on 11 March 2014).

NEDICC. 2016. ***Invitation to first Library Carpentry Workshop in Africa, 25-26 August 2016***. [Sl.]: NEDICC in collaboration with North-West University and Talarify. [Unpublished].

NELSON MANDELA UNIVERSITY. 2017. ***Centre for Broadband Communication.*** Port Elizabeth: Nelson Mandela University. [Online] available at http://broadband.mandela.ac.za/ (Accessed 14 September 2017).

NENTWICH, M. 2003. ***Cyberscience: research in the age of the Internet.*** Vienna: Austrian Academy of Sciences.

NETHERLANDS ESCIENCE CENTER (NLESC). 2015. ***Strategy 2015-2020 and beyond.*** Netherlands eScience Center (NLeSC). [Online] available at https://www.esciencecenter.nl/img/pressroom/331-010_ESC_Strategy_Brochure_LR_spreads.pdf (Accessed 21 September 2017).

NETHERLANDS ORGANISATION FOR SCIENTIFIC RESEARCH (NWO). 2017. ***South African-Dutch data science partnership to address SKA big data challenge, 15 December 2016.*** Netherlands Organisation for Scientific Research (NWO). The Hague; Utrecht, Netherlands: Netherlands Organisation for Scientific Research. [Online] available at https://www.nwo.nl/en/news-and-events/news/2016/ew/south-african-dutch-data-science-partnership-to-address-ska-big-data-challenge.html (Accessed 14 September 2017).

NEUMAN, W.L. 2000. ***Social research methods: qualitative and quantitative approaches.*** 4th ed. Boston: Allyn and Bacon.

NEUROTH, H., LOHMEYER, F. & SMITH, K.M. 2011. TextGrid-Virtual Research Environment for the Humanities. ***International Journal of Digital Curation***, 6(2): 223-331. [Online] available at http://www.ijdc.net/index.php/ijdc/article/view/193 (Accessed 18 April 2015).

New MPhil (Specialisation in Digital Curation), 2014. Cape Town: Libraries and Information Studies Centre, University of Cape Town (UCT). [Online] available at http://www.lib.uct.ac.za/lisc/2014/05/14/new-mphil-specialisation-in-digital-curation/ (Accessed 7 July 2014).

New national digital repository for social and economic data. Manchester, UK: University of Manchester, 2012a. [Online] available at http://www.manchester.ac.uk/discover/news/article/?id=8546 (Accessed 26 November 2016).

New national digital repository for social and economic data, 2012b**.** ESRC news release, 24 July 2012. Swindon, UK: ESRC. [Online] available at https://web.archive.org/web/20120807051248/http://www.esrc.ac.uk/news-and-events/press-releases/22231/new-national-digital-repository-for-social-and-economic-data.aspx (Accessed 1 December 2016).

The new Oxford style manual. 2nd.ed. Edited by Robert M. Ritter. Oxford University Press, 2012.

NGUYEN, N., WEGENER, M. & RUSSEL, I. 2006. Risk management tools for dryland farmers in southwest Queensland: an action research approach. In: ***Proceedings of APEN International Conference 2006***, held 6-8 March 2006, at Beechworth, Victoria, Australia. [Online] available at http://www.regional.org.au/au/apen/2006/refereed/1/2927_nguyenn.htm#TopOfPage (Accessed 9 February 2017).

NIELSEN, H.J. & HJØRLAND, B. 2014. Curating research data: the potential roles of libraries and information professionals, ***Journal of Documentation***, 70(2): 221-240.

NIELSEN, J. 1993. Iterative user-interface design. ***IEEE Computer,*** November, 26(11): 32-41.

NORMAN, B. & STANTON, K.V. 2014. From project to strategic vision: taking the lead in research data management support at the University of Sydney Library. ***International Journal of Digital Curation***, 9(1): 253-262. [Online] available at http://www.ijdc.net/index.php/ijdc/article/view/316 (Accessed 12 September 2017).

NORMANDIEU, K. 2013. Beyond volume, variety and velocity is the Issue of big data veracity. ***Inside Big Data***, 12 September 2013. [Online] available at http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/ (Accessed 25 September 2014).

NORTHERN ILLINOIS UNIVERSITY. n.d. ***Responsible conduct in data management.*** DeKalb, IL: Faculty development and instructional Design Center, Northern Illinois University. [Online] available at https://web.archive.org/web/20170122082405/https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html (Accessed 20 January 2017).

NRF. n.d. ***Digitisation and Data Preservation Centre.*** Pretoria: National Research Foundation. [Online] available at http://digi.nrf.ac.za/ (Accessed 6 February 2017).

NSF. 2007. ***Sustainable Digital Data Preservation and Access Network Partners (DataNet): Program Solicitation.*** (NSF 07-601). Arlington, VA: National Science Foundation, Office of Cyberinfrastructure, Directorate for Computer and Information Science and Engineering. [Online] available at https://www.nsf.gov/pubs/2007/nsf07601/nsf07601.pdf (Accessed 2 September 2017).

NSF. 2009. ***DataNet full proposal: the Data Conservancy: a digital research curation virtual organization: Award Abstract #0830976.*** Arlington, VA: National Science Foundation, Office of Cyberinfrastructure, Directorate for Computer and Information Science and Engineering. [Online] available at https://www.nsf.gov/awardsearch/ (Accessed 2 September 2017).

NSF. 2014. ***DataONE: Data Observation Network for Earth: Award Abstract #1430508.*** Arlington, VA: National Science Foundation, Office of Cyberinfrastructure, Directorate for Computer and Information Science and Engineering. [Online] available at https://www.nsf.gov/awardsearch/ (Accessed 2 September 2017).

OASIS reference model for Service Oriented Architecture 1.0, 2006. Edited by C. Matthew MacKenzie, Ken Laskey, Francis McCabe, Peter F. Brown, and Rebekah Metz. [Sl.]: OASIS Open. [Online] available at http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf (Accessed 29 January 2012).

O'BRIEN, L. 2005. E-Research: an imperative for strengthening institutional partnerships. ***Educause Review***, November / December, 40(6): 64-76.

OCHEM, N. 2008. ***Business logic management in a web application.*** Masters Thesis, Royal Institute of Technology, Stockholm, Sweden. Stockholm.

OCHIENG, O.B. 2016. Data-intensive research capacity boosted ahead of SKA. ***University World News***, Global Edition, 23 September, Issue 429.

ONWUEGBUZIE, A.J., LEECH, N.L. & COLLINS, K.M.T. 2012. Qualitative analysis techniques for the review of the literature. ***The Qualitative Report***, 17(56): 1-28. [Online] available at http://www.nova.edu/ssss/QR/QR17/onwuegbuzie.pdf (Accessed 17 September 2014).

OPENAIRE. 2016. ***What is the Open Research Data Pilot?*** [Sl.]: OpenAIRE. [Online] available at https://www.openaire.eu/opendatapilot (Accessed 19 September 2017).

OpenWetWare, 2009. Cambridge, MA: BioBricks Foundation. [Online] available at http://openwetware.org/wiki/Main_Page (Accessed 25 September 2012).

O'REILLY, T. 2005. What is Web 2.0: design patterns and business models for the next generation of software. O'Reilly, web log post 30 September 2005. Sebastopol, CA: O'Reilly Media. [Online] available at http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1 (Accessed 12 February 2018).

OWEN, J.M. & ROGERS, P.J. 1999. ***Program evaluation: forms and approaches.*** 2nd ed. St Leonards, NSW, Australia: Allen and Unwin.

The Oxford Dictionary. [sl.]: Oxford University Press, 2014. [Online] available at http://www.oxforddictionaries.com/us/ (Accessed 16 September 2014).

PAGE-SHIPP, R., HAMMES, M.M.P., PIENAAR, H., REAGON, F., THOMAS, G., VAN DEVENTER, M.J. & VELDSMAN, S. 2005. eResearch: the challenge for South Africa. South African ***Journal of Information Management***, December, 7(4). [Online] available at http://www.sajim.co.za/index.php/SAJIM/article/viewFile/287/277 (Accessed 16 September 2017).

Participatory action research. Edited by William Foote Whyte. (Sage focus edition; vol. 123). Newbury Park, CA: Sage Publications.

PARTNERSHIP FOR ACCESSING DATA IN EUROPE (PARADE). 2009***. Strategy for a European Data Infrastructure: White Paper.*** [Sl.]: Partnership for Accessing Data in Europe (PARADE). [Online] available at https://web.archive.org/web/20111112164037/http://www.csc.fi/english/pages/parade/whitepaper (Accessed 19 September 2017).

PATERSON, M., LINDSAY, D., MONOTTI, A. & CHIN, A. 2007. DART: a new missile in Australia's e-research strategy. ***Online Information Review***, 31(2): 116-134.

PATTERTON, L. 2016. ***RDM practices at the CSIR: results of two studies.*** Presented at the NeDiCC meeting, 18 February, CSIR, Pretoria. [Unpublished].

PAULOVICH, B. 2015. Design to improve the health education experience: using participatory design methods in hospitals with clinicians and patients. ***Visible Language Journal,*** April, 49(1-2). [Online] available at http://visiblelanguagejournal.com/issue/161/article/971 (Accessed 17 January 2017).

PAZ, V. 2015. ***RDSI: changing the face of research data storage in Australia.*** Presentation delivered at the Terena Networking Conference 2015, held in Porto, Portugal, 15-18 June 2015. [Online] available at https://tnc15.terena.org/getfile/2803 (Accessed 4 January 2017).

PAZ, V. & TATE, N. 2014. Accessible research storage: RDSI project transitioning and lessons learned along the way. Presented at the ***eResearch Australasia 2014 Conference***, 27-31 October, Melbourne, Auatralia. [Online] available at https://eresearchau.files.wordpress.com/2014/07/eresau2014_submission_38.pdf (Accessed 4 January 2017).

PENN STATE UNIVERSITY LIBRARIES. 2014. ***What is data management?*** State College, PA: Pennsylvania State University Libraries. [Online] available at http://www.libraries.psu.edu/psul/pubcur/what_is_dm.html (Accessed 10 March 2014).

PERRY, S. 2013. SKA: the ultimate big data challenge. ***Brainstorm Magazine***, 2 May. [Online] available at http://www.brainstormmag.co.za/index.php?option=com_content&view=article&id=4900:ska-the-ultimate-big-data-challenge (Accessed on 5 March 2014).

PETERS, D. 2013. DIRISA: a roadmap for South African data research infrastructures. Presented at the ***eResearch Africa 2013 Conference***, 6-10 October, Newlands, Cape Town. [Online] available at: http://eresearch.ac.za/wp-content/uploads/2013/05/DIRISA_Roadmap.pdf (Accessed 29 April 2014).

PHAM, T.V., LAU, L.M.S., DEW, P.M., & PILLING, M.J. 2005. Collaborative e-Science architecture for Reaction Kinetics Research Community. In: ***Proceedings of Challenges of large applications in distributed environments, 2005***. [Sl.]: IEEE, p.13-22.

PICKARD, A.J. 2007. ***Research methods in information.*** London: Facet Publishing.

PIENAAR, H. 2010. ***Findings of a survey of research data management practices over the period October 2009-March 2010 at the University of Pretoria (UP): undertaken by the Department of Library Services in order to improve research practices.*** Pretoria: Department of Library Services, University of Pretoria, 2010.

PIENAAR, H. 2017. ***Goeie nuus!*** E-mail received on 24 August 2017.

PIENAAR, H. & VAN DEVENTER, M. 2009. To VRE or not to VRE: do South African Malaria researchers need a Virtual Research Environment? ***Ariadne***, 30 April, Issue 59. [Online] available at http://www.ariadne.ac.uk/print/issue59/pienaar-vandeventer (Accessed 6 August 2012).

PLÖSCH, R. 2004. ***Contracts, scenarios and prototypes: an integrated approach to high quality software.*** Berlin; Heidelberg: Springer.

POTHEN, P. 2004. ***Developing the UK's e-infrastructure for science and innovation: report of the OSI e-infrastructure Working Group.*** [S.l.]: OSI e-infrastructure Working Group. [Online] available at https://web.archive.org/web/20160320220924/http://www.nesc.ac.uk/documents/OSI/report.pdf (Accessed 12 February 2018).

Preservation, 2012. Cambridge, UK: University Library and the University Computing Service, the University of Cambridge. [Online] available at http://www.lib.cam.ac.uk/dataman/pages/preservation.html (Accessed 19 September 2014).

Preservation definitions, 2011. ESIP Federation Wiki, wiki article. [Online] available at http://wiki.esipfed.org/index.php/Preservation_Definitions (Accessed 19 August 2014).

Preserving and providing access to South African social science and humanities research data: science seminar, 2012. Hosted by the Department of Science and Technology (DST), the University of Pretoria (UP), and the Human Sciences Research Council (HSRC), 5 November 2012, CSIR Convention Centre. [Unpublished].

Prometheus, n.d*.* Köln, Germany: Kunsthistorisches Institut der Universität zu Köln [Online] available at http://prometheus-bildarchiv.de/ (Accessed 9 August 2012).

PRYOR, G. 2014. A patchwork of change. In: ***Delivering research data services: fundamentals of good practice.*** Edited by Graham Pryor, Sarah Jones and Angus Whyte. London: Facet Publishing, p.1-19.

PUDER, A., RÖMER, K. & PILHOFER, F. 2006. ***Distributed systems architecture***. San Francisco, CA: Morgan Kaufman.

PURDUE UNIVERSITY. 2011. ***Latest release makes HUBzero computing, collaboration software more social***, 4 April 2011. West Lafayette, IN: Purdue University. [Online] available at http://www.rcac.purdue.edu/news/detail.cfm?newsId=459 (Accessed 24 March 2013).

PURDUE UNIVERSITY. 2013. ***Purdue University Research Repository.*** West-Lafayette, IN: Purdue University. [Online] available at https://purr.purdue.edu/ (Accessed 11 March 2014).

Quantitative data: surveys, 2014. From Data Collection: Primary Research Methods Tutorial. KnowThis.com. [Online] available at http://www.knowthis.com/data-collection-primary-research-methods/quantitative-data-surveys (Accessed 16 September 2014)

QUINT, B. 2004. OECD Ministers support Open Access for publicly funded research data***. Information Today-Newsbreaks***, 9 February 2004. [Online] available at: http://newsbreaks.infotoday.com/NewsBreaks/OECD-Ministers-Support-Open-Access-for-Publicly-Funded-Research-Data-16519.asp (Accessed on 9 April 2014).

RAGHUNATHAN, B. 2013. ***The complete book of data anonymization: from planning to implementation.*** Broken Sound Parkway, NW: CRC Press, Taylor and Francis Group.

RESEARCH COUNCILS UK. 2014. ***RCUK Common Principles on Data Policy***. Swindon, UK: Research Councils UK. [Online] available at http://www.rcuk.ac.uk/research/datapolicy/ (Accessed 4 December 2016).

RDSI project 2010-2015. [Sl.]: National Computational Infrastructure, 2015. [Online] available at http://nci.org.au/about-nci/history/rdsi-project-2010-2015/ (Accessed 4 January 2017).

RefWorks, 2009. Cambridge, UK: ProQuest. [Online] available at http://www.refworks.com/ (Accessed 15 October 2012).

Research Data Alliance. n.d.(a). [Online] available at https://rd-alliance.org/ (Accessed 5 March 2014).

RESEARCH DATA ALLIANCE. n.d.(b). ***European Data Infrastructure (EUDAT).*** [Sl.]: Research Data Alliance. [Online] available at http://rd-alliance.org/european-data-infrastructure-eudat.html (Accessed 18 August 2016).

Research Data Australia, n.d. [Online] available at https://researchdata.ands.org.au (Accessed 12 september 2017).

Research data management: principles, practices and prospects. Washington, DC: DataRes, Council on Library and Information Resources, 2013. [Online] available at http://www.clir.org/pubs/reports/pub160/pub160.pdf (Accessed 19 March 2014).

Research Data Services (ReDS), 2014. [Sl.]: University of Queensland, RDSI. [Online] available at https://web.archive.org/web/20150317023308/https://www.rdsi.edu.au/reds (Accessed 4 January 2017).

RESEARCH DATA SERVICES, UNIVERSITY OF WISCONSIN-MADISON.2014. ***Data storage and backup.*** Madison, WI: Research Data Services, University of Wisconsin-Madison. [Online] available at http://researchdata.wisc.edu/manage-your-data/data-backup-and-integrity/ (Accessed 19 September 2014).

ResearchGate, 2012***.*** Cambridge, MA: ResearchGate Corporation; Berlin, Germany: ResearchGate GmbH. [Online] available at http://www.researchgate.net/ (Accessed 4 October 2012).

Research Information Centre (RIC), n.d. London, UK: The British Library. [Online] available at http://www.bl.uk/reshelp/experthelp/science/ric/ric.html (Accessed 22 March 2013)

Research Information Centre Framework, 2012. [Sl.]: Microsoft Research. [Online] available at http://research.microsoft.com/en-us/projects/ric/ (Accessed 9 August 2012).

Research use: HSRC, 2014. Pretoria: HSRC. [Online] available at http://www.hsrc.ac.za/en/ria/about-us/research-use (Accessed on 13 March 2014).

Roadmap: global research data management advisory platform combines DMPTool and DMPOnline. ***Science Codex***, 31 March, 2016. [Online] available at http://www.sciencecodex.com/roadmap_global_research_data_management_advisory_platform_combines_dmptool_and_dmponline-179054 (Accessed 30 August 2017).

ROBERTSON LIBRARY. n.d. ***VREs (Virtual Research Environments).*** Charlotteville, PEI: Robertson Library, University of Prince Edward Island.

ROBSON, C. 1993. ***Real world research.*** Oxford: Blackwell.

ROBSON, C. 1997***. Real world research: a resource for social scientists and practitioner-researchers.*** Oxford: Blackwell.

ROBSON, C. 2011. ***Real world research: a resource for users of social research methods in applied settings.*** Chichester, West Sussex: Wiley.

ROBSON, R. 1999. WWW-based course-support systems: the first generation. ***International Journal of Educational Telecommunications***, 5(4): 271-282.

ROOS, K. & MIAS, E. 2017. Suggesting an institutional data repository for UCT. Presented at the ***eResearch Africa 2017 Conference***, 2-5 May, University of Cape Town, Cape Town, South Africa. [Online] available at http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/2017-05-04_UCT-IDR_eRA2017_abridged.pdf (Accessed 16 September 2017).

ROOS, K., MIAS, E. & VAN ROOYEN, J. 2017. ***Suggesting an institutional data repository for UCT: eResearch Stakeholder Initiative.*** Presented online at the ***NeDICC meeting***, 26 January, CSIR, Pretoria, South Africa. [Unpublished].

ROSE, D. 2007. Application. ***Techtarget.*** [Online] available at http://searchsoftwarequality.techtarget.com/definition/application (Accessed 12 February 2018).

ROSENBAUM, S. 2010. Data Governance and Stewardship: designing data stewardship entities and advancing data access. ***Health Services Research***, October, 45(5), Part 2: 1442–1455.

ROUSE, M. 2007. Data governance. ***Techtarget.*** [Online] available at http://searchdatamanagement.techtarget.com/definition/data-governance (Accessed 18 August 2014).

ROUSE, M. 2010. Data archiving. ***Techtarget.*** [Online] available at http://search databackup.techtarget.com/definition/data-archiving (Accessed 19 August 2014).

ROUSE, M. & BIGELOW, S.J. 2017. Cloud computing. ***Techtarget.*** [Online] available at http://searchcloudcomputing.techtarget.com/definition/cloud-computing (Accessed 20 November 2017).

ROUSE, M., CHURCHVILLE, F. & DANG, M. 2016. User interface (UI). ***Techtarget.*** [Online] available at http://searchmicroservices.techtarget.com/definition/user-interface-UI (Accessed 12 February 2018).

THE ROYAL SOCIETY. 2016. ***Costs of digital repositories.*** [Sl.]: The Royal Society, 2016. [Online] available at https://royalsociety.org/topics-policy/projects/science-public-enterprise/digital-repositories/ (Accessed 25 November 2016).

RUDESTAM, K.E. & NEWTON, R.R. 1992. ***Surviving your dissertation.*** London: Sage.

RULE, P. & VAUGHN, J. 2011. ***Your guide to case study research.*** Hatfield, Pretoria: Van Schaik.

RUMSEY, D.J. 2015. Types of statistical data: numerical, categorical, and ordinal. ***Statistics for Dummies.*** [Online] available at http://www.dummies.com/how-to/content/types-of-statistical-data-numerical-categorical-an.html (Accessed 9 May 2015).

RUNESON, P. & HÖST, M. 2009. Guidelines for conducting and reporting case study research in software engineering. ***Empirical Software Engineering***, April, 14(2): 131-164.

RUNESON, P., HÖST, M., RAINER, A. & REGNELL, B. 2012. ***Case study research in software engineering: guidelines and examples.*** Hoboken, N.J.: John Wiley and Sons.

SABIF: The South African Biodiversity Information Facility, 2014. [Sl.]: SABIF. [Online] available at http://www.sabif.ac.za/ (Accessed 28 April 2014).

Sakai, n.d. [Sl.: Sakai Foundation]. [Online] available at http://www.sakaiproject.org/ (Accessed 25 September 2012).

SA National Committee for CODATA: overview, n.d. Pretoria, National Research Foundation. [Online] available at http://www.nrf.ac.za/sites/default/files/documents/CODATA%20profile_SA%20ICSU%20v1.pdf (Accessed 4 February 2017).

SANDLAND, R. 2009. Introduction to ANDS. ***Share: Newsletter of the Australian National Data Service***, July, 1(1). [Online] available at http://www.ands.org.au/__data/assets/pdf_file/0005/388841/share-issue-1.pdf (Accessed 3 January 2017).

SANPARKS. n.d. ***South African National Parks Data Repository.*** [Pretoria]: SANParks. [Online] available at http://dataknp.sanparks.org/sanparks/style/skins/sanparks/ (Accessed 14 September 2017).

SANPARKS. 2016. ***SANParks Annual Report 2015/16.*** (RP98/2016). [Pretoria]: SANParks. [Online] available at https://www.sanparks.org/assets/docs/general/annual-report-2016.pdf (Accessed 14 September 2017).

SANPARKS. 2017. ***About us.*** [Pretoria]: SANParks. [Online] available at https://www.sanparks.org/about/ (Accessed 14 September 2017).

SANReN: South African National Research Network, n.d. [Sl.]: SANReN [Online] available at http://www.sanren.ac.za/overview/ (Accessed 7 July 2014).

SARIMA: Southern African Research and Innovation Management Association, 2014. Pretoria, SARIMA. [Online] available at https://web.archive.org/web/20140625075339/http://www.sarima.co.za/about-us/ (Accessed 3 December 2016).

SARIMA: Southern African Research and Innovation Management Association, 2016. Pretoria, SARIMA. [Online] available at http://www.sarima.co.za (Accessed 3 December 2016).

SCHIZOPHRENIA RESEARCH FORUM. 2017a. ***History.*** [Sl.]: Schizophrenia Research Forum. [Online] available at https://www.schizophreniaforum.org/history (Accessed 22 September 2017).

SCHIZOPHRENIA RESEARCH FORUM. 2017b. ***Mission.*** [Sl.]: Schizophrenia Research Forum. [Online] available at https://www.schizophreniaforum.org/mission (Accessed 22 September 2017).

SCHNELL, K. & SHETTERLEY, N. 2013. ***Understanding data visualization.*** [Sl.]: Accenture. [Online] available at http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Tech-Labs-Data-Visualization-Full-Paper.pdf (Accessed 19 September 2014).

SCHROTH, C. & JANNER, T. 2007. Web 2.0 and SOA: converging concepts: enabling the Internet of services. ***IT Professional***, May / June, 9(3): 36-41.

SCHURINK, E.M. 1998. The methodology of unstructured face-to-face interviewing. In: De Vos, A.S. & H Strydom, ***Research at grass roots: a primer for the caring professions.*** Pretoria: Van Schaik, p. 297-300.

Science Gateways Community Institute, n.d. [Online] available at https://sciencegateways.org/ (Accessed 23 September 2017).

SCIENCE INTERNATIONAL. 2015. ***Open data in a big data world.*** Paris: International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership (IAP). [Online] available at https://icsu.org/cms/2017/04/open-data-in-big-data-world_long.pdf (Accessed on 19 September 2017).

SciVerse Scopus, 2012. Amsterdam, Netherlands: Elsevier B.V. [Online] available at http://www.info.sciverse.com/scopus (Accessed 20 October 2012).

Scratchpads: biodiversity online, n.d. [Sl.]: European Distributed Institute of Taxonomy. [Online] available at http://scratchpads.eu/ (Accessed 6 October 2012).

SCRIVEN, M. 1991. ***Evaluation thesaurus.*** 4th ed. Newbury Park, CA: Sage.

SEAD, n.d. [Online] available at http://sead-data.net/ (Accessed 2 September 2017).

SEARIGHT, H.R. et al. 2011. E-Research in the social sciences: the possibilities and the reality: a review article. ***Current Research Journal of Social Sciences***, 30 March, 3(2): 71-80.

SEARLE, S., WOLSKI, M., SIMONS, N. & RICHARDSON, J. 2015. Librarians as partners in research data service development at Griffith University. ***Program: Electronic library and information systems***, 40(4): 440-460.

SERGEANT, D.M., ANDREWS, S. & FARQUHAR, A. 2006. ***Embedding a VRE in an institutional environment EVIE: workpackage 2: user requirements analysis.*** (University of Leeds technical report). [Sl.: University of Leeds] [Online] available at https://www.leeds.ac.uk/evie/workpackages/wp2/evieWP2_UserRequirementsAnalysis_v1_0.pdf (Accessed 2 September 2012).

Shibboleth, 2012. Ann Arbor, MI; Washington, DC: Internet2. [Online] available at http://shibboleth.net/. (Accessed 25 September 2012).

SIM, Y.W., WANG, C., GILBERT, L. & WILLS, G.B. 2005. ***An Overview of Service-Oriented Architecture.*** (Technical Report; ECSTR-IAM05-004). Southampton: University of Southampton, p.1-8. [Online] available at http://eprints.soton.ac.uk/261209/ (Accessed 3 March 2015).

SIMEONI, F., PAGANO, P., SIMI, M. & CONNOR, R. 2008. Application-level research e-infrastructures: the gCube approach. ***UK e-Science All Hands Meeting 2008***, 8-11 September, Edinburgh, UK, [Online] available at https://www.researchgate.net/profile/Donatella_Castelli/publication/237797452_Application-level_Research_e-Infrastructures_the_gCube_Approach/links/0deec528127e32b71e000000.pdf?inViewer=0&pdfJsDownload=0&origin=publication_detail (Accessed 25 September 2017).

SIMMS, S., JONES, S., ASHLEY, K., RIBEIRO, M., CHODACKI, S., ABRAMS, S. & STRONG, M. 2016. Roadmap: a research data management advisory platform. ***Research Ideas and Outcomes***, 30 March, 2(e8649). [Online] available at http://riojournal.com/articles.php?id=8649 (Accessed 13 December 2016).

SIMPSON, J. n.d. Data Masking and Encryption Are Different. ***IRI Blog Articles***, web log post. [Online] available at http://www.iri.com/blog/data-protection/data-masking-and-data-encryption-are-not-the-same-things/ (Accessed 18 September 2014).

SINGH, M. & VOUK, M. 1996. Scientific Workflows: scientific computing meets transactional workflows. In: ***Proceedings of the NSF Workshop on Workflow and Process Automation in Information Systems: State-of-the-art and Future Directions***. Edited by A. Sheth, Athens, GA, 8-10 May, 1996. [Online] available https://www.csc2.ncsu.edu/faculty/mpsingh/papers/databases/workflows/sciworkflows.html (Accessed 12 February 2018).

SKA SOUTH AFRICA (a). n.d. ***The project.*** Cape Town; Johannesburg: SKA South Africa. [Online] available at http://www.ska.ac.za/about/the-project/ (Accessed 13 September 2017).

SKA SOUTH AFRICA (b). n.d. ***MeerKAT radio telescope.*** Cape Town; Johannesburg: SKA South Africa. [Online] available at http://www.ska.ac.za/gallery/meerkat/ (Accessed 13 September 2017).

SKA SOUTH AFRICA. 2015. ***Minister opens the way for big data: media release.*** Pinelands: SKA South Africa. [Online] available at http://ska.ac.za/releases/20150329.php (Accessed 4 September 2015).

Skype, 2012. Luxembourg: Skype Software S.à r.l / Skype Communications S.à r.l. [Online] available at http://www.skype.com (Accessed 15 October 2012).

SlideShare, 2012. San Francisco, CA: SlideShare Inc. [Online] available at http://www.slideshare.com (Accessed 20 October 2012).

SMIT, J. 2015. ***SKA is helping to bring Africa into the knowledge economy.*** Pretoria: University of Pretoria. [Online] available at http://www.up.ac.za/en/faculty-of-engineering-built-environment-it/news/post_2101647-ska-help-om-afrika-die-kennisekonomie-binne-te-lei (Accessed 3 December 2016).

SMITH, J. 2012. ***Evaluating training - what you need to know: definitions, best practices, benefits and practical solutions.*** Dayboro, Australia: Emereo Publishing.

Social construction: a reader. Edited by Mary Gergen and Kenneth J. Gergen. London: Sage, 2003.

Sol Plaatje University Annual Report 2015. Edited by Hollie Clarkson, Annemarie van der Nest and Marietjie Grobbelaar. Kimberley, Office of the Registrar, Sol Plaatje University, 2015. [Online] available at http://www.spu.ac.za/docs/spu_annual_2015.pdf (Accessed 4 January 2017).

SOUTH AFRICA. 2008. Human Sciences Research Council Act (No. 17 of 2008). ***Government Gazette***, 30 September 2008, 519(31470). Cape Town: The Presidency.

South Africa Yearbook 2015/16. 23nd ed. Edited by Elias Tibane and Nomfundo Lentsoane. Pretoria: Government Communications (GCIS), 2016. [Online] available at http://www.gcis.gov.za/content/resourcecentre/sa-info/yearbook2015-16 (Accessed 2 February 2017).

South African Astroinformatics Alliance (SA³), 2014. [Online] available at http://www.sa3.ac.za/ (Accessed 28 April 2014).

South African Data Archive, n.d. [Online] available at http://sada.nrf.ac.za/introduction.html (Accessed 30 July 2014)

South African National Grid, n.d*.* [Online] available at http://www.sagrid.ac.za/ (Accessed 28 April 2014).

SPIRO, L. 2009. Examples of collaborative digital humanities projects: facilitating communication and knowledge building: collaboratories. ***Digital Scholarship in the Humanities***, web log post, 1 June 2009*.* [Online] available at https://digitalscholarship.wordpress.com/2009/06/01/examples-of-collaborative-digital-humanities-projects/ (Accessed 5 February 2012).

STERN, P. & PORR, C. 2011. ***Essentials of accessible grounded theory.*** Walnut Creek, California: Left Coast Press.

STEWART, C.A. 2010. What is Cyberinfrastructure? Presented at the ***ACM SIGUCCS 2010 Annual Meting***, 24-27 October, Norfolk, VA. [Online] available at http://hdl.handle.net/2022/13987 (Accessed 20 February 2015).

STRASSER, C., COOK, R., MICHENER, W. & BUDDEN, A. 2012. ***Primer on data management: what you always wanted to know.*** [Albuquerque, NM]: DataONE, [University of New Mexico], p.1-11. [Online] available at http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf (Accessed 28 August 2013)

STREUBERT, H.J. & CARPENTER, D.R. 1995. ***Qualitative research in nursing: advancing the humanistic imperative.*** Philadelphia, PA: J. B. Lippincott Company.

SURFFOUNDATION. n.d. ***Collaboratories.*** [Sl.]: SURFfoundation. [Online] available at http://www.surffoundation.nl/en/themas/openonderzoek/collaboratories/pages/default.aspx (Accessed 26 February 2012).

SWANBORN, P.G. 2010. ***Case study research: what, why and how?*** London: Sage.

TANG, H., WU, Y. WANG, G. & YAO, Y.Y. 2003. CUPTRSS: a Web-based Research Support System. In: ***Proceedings of the Workshop on Applications Products and Services of Web-based Support Systems WSS03.*** Edited by J.T. Yao, and P. Lingras. Halifax, Canada, p. 21-28. [Online] available at http://www2.cs.uregina.ca/~wss/wss03/03/wss03-21.pdf (Accessed 11 February 2012).

TAYLOR, R. 2016. ***African Research Cloud Workshop.*** Held 27-28 October, Pretoria, South Africa, 27-28. [Sl.]: Inter-University Institute for Data Intensive Astronomy. [Online] available at http://idia.ac.za/ARCworkshop (Accessed 16 November 2016).

Taverna, 2009. Manchester, UK: School of Computer Science, University of Manchester. [Online] available at http://www.taverna.org.uk/ (Accessed 25 September 2012).

TENOPIR, C., TALJA, S., HORSTMANN, W., LATE, E., HUGHES, D., POLLOCK, D., SCHMIDT, B., BAIRD, L., SANDUSKY, R. & ALLARD, S. 2017. Research data services in European Academic Research Libraries. ***Liber Quarterly***, 27(1): 23-44. [Online] available at https://www.liberquarterly.eu/articles/10.18352/lq.10180/ (Accessed 19 September 2017).

TEXAS A & M UNIVERSITY LIBRARIES. n.d. ***Research Data Management.*** (Research Guides). College Station: Texas A & M University Libraries. [Online] available at http://guides.library.tamu.edu/DataManagement (Accessed 19 August 2014).

TextGrid, 2012. Göttingen, Germany: [German Federal Ministry of Education and Research (BMBF)]. [Online] available at http://www.textgrid.de/en/ (Accessed 6 February 2013).

THANOS, C. 2013. A vision for global research data infrastructures. ***Data Science Journal***, 13 September, 12: 71-90.

TIZARD, J. 2014. NRN project launch. ***National Research Project News***, 28 November. [Online] available at http://www.nrn.edu.au/news/nrnprojectlaunch (Accessed 4 January 2017).

Thomson Reuters (ISI) Web of Knowledge, 2012. New York, NY: Thomson Reuters. [Online] available at http://thomsonreuters.com/products_services/science/science_products/a-z/isi_web_of_knowledge/ (Accessed 20 October 2012).

The R project for statistical computing, 2012***.*** Boston, MA: Free Software Foundation. [Online] available at http://www.r-project.org/ (Accessed 16 October 2012).

TRELOAR, A. 2009. Design and implementation of the Australian National Data Service. ***The International Journal of Digital Curation***, 4(1): 125-137.

TRELOAR, A., CHOUDHURY, G.S. & MICHENER, W. 2012. Contrasting national research data strategies: Australia and the USA. In: ***Managing research data.*** Edited by Graham Pryor. London: Facet Publishing, p.173-203.

Triana, 2012. Cardiff, UK: Cardiff University. [Online] available at http://www.trianacode.org/ (Accessed 4 October 2012).

UCD LIBRARY. 2014***. Research Data Management: why mange research data?*** Dublin, Republic of Ireland: University College Dublin. [Online] available at http://libguides.ucd.ie/data/why_manage (Accessed 21 October 2014).

UCLA LIBRARY. 2014. ***Data Management for the Humanities, UCLA Library***. Los Angeles, CA: University of California, Los Angeles, 2014. http://guides.library.ucla.edu/content.php?pid=440928&sid=3632834 (Accessed 17 Aug 2014).

UC SAN DIEGO. 2014. ***Integrated Digital Infrastructure: data curation.*** San Diego, CA: UC San Diego. [Online] available at http://idi.ucsd.edu/data-curation/definition.html (Accessed 25 October 2014).

UK DATA ARCHIVE. 2007a. ***Across the decades: 40 years of data archiving.*** Colchester: UK Data Archive, 2007. [Online] available at http://www.data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf (Accessed 5 March 2014).

UK DATA ARCHIVE. 2007b. ***40th UKDA.*** Essex, UK: UK Data Archive, University of Essex, 2007. [Online] available at http://ukda40.data-archive.ac.uk/about/Introduction .asp (Accessed 26 Ovember 2016)

UK DATA ARCHIVE. 2014. ***Research data lifecycle.*** Colchester, UK: UK Data Archive, University of Essex. [Online] available at http://www.data-archive.ac.uk/create-manage/life-cycle (Accessed 25 October 2014).

UK DATA ARCHIVE. 2016. ***Our projects.*** Colchester, UK: UK Data Archive, University of Essex. [Online] available at http://www.data-archive.ac.uk/about/projects/cessda-elsst (Accessed 26 November 2016).

UK DATA SERVICE. 2013. ***Data management costing tool.*** Colchester, UK: UK Data Archive, University of Essex. [Online] available at http://www.data-archive.ac.uk/media/247429/costingtool.pdf (Accessed on 1 December 2016).

UK DATA SERVICE. 2016. ***Our stakeholders.*** Colchester, UK: UK Data Service. [Online] available at https://www.ukdataservice.ac.uk/about-us/stakeholders (Accessed 30 November 2016).

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY (US EPA). 2002. ***Guidance on Environmental Data Verification and Data Validation: EPA QA/G-8.*** Washington, DC: Environmental Protection Agency. [Online] available at http://www.epa.gov/QUALITY/qs-docs/g8-final.pdf (Accessed 24 September 2014).

UNITED STATES OF AMERICA. 1996. ***Health Insurance Portability and Accountability Act of 1996.*** Public Law 104-191, 104th Congress [H.R. 3103]. [Washington, DC]: United States of America. [Online] available at https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf (Accessed 31 March 2017.

UNITED STATES WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY. 2013. ***Increasing access to the results of Federally funded scientific research.*** (Memorandum for the heads of executive departments and agencies) [Online] available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (Accessed 3 September 2017).

UNIVERSIDAD DE CORDOBA, n.d. ***Access Grid.*** Cordoba, Spain: Universidad de Cordoba. [Online] available at http://www.accessgrid.org/ (Accessed 21 February 2015).

UNIVERSITY OF CAPE TOWN LIBRARIES. 2017. ***Digital Library Services.*** Rondebosch, Cape Town: University of Cape Town Libraries. [Online] available at http://www.digitalservices.lib.uct.ac.za/ (Access 5 January 2017).

UNIVERSITY OF EDINBURGH. 2014. ***Why manage research data?*** Edinburgh, UK: University of Edinburgh. [Online] available at http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/research-data-mgmt/why-manage (Accessed 21 October 2014).

UNIVERSITY OF MANCHESTER LIBRARY. n.d. ***Why you should manage your data.*** Manchester. UK: University of Manchester Library. [Online] available at http://www.library.manchester.ac.uk/ourservices/research-services/rdm/whyyoushouldmanageyourdata/ (Accessed 21 October 2014).

UNIVERSITY OF MINESOTA LIBRARIES. 2014. ***What is Data Management?*** Minneapolis, MN: University of Minnesota. [Online] available at: https://www.lib.umn.edu/datamanagement/whatdata (Accessed 17 Aug 2014).

UNIVERSITY OF TENESSEE LIBRARIES. 2014. ***Data Management.*** (Research Guides). Knoxville, TN: University of Tenessee Libraries. [Online] available at http://libguides.utk.edu/datamanagement (Accessed 19 August 2014).

UNIVERSITY OF VIRGINIA LIBRARY. 2014. ***Why manage your data?*** Charlottesville, VA: University of Virginia. [Online] available at http://dmconsult.library.virginia.edu/whymanage/ (Accessed 21 October 2014).

UNIVERSITY OF THE WITWATERSRAND. 2017a. ***BSc Honours in Big Data Analytics.*** Johannesburg: University of the Witwatersrand. [Online] available at https://www.wits.ac.za/csam/computer-science/postgraduate/big-data-analytics/ (Accessed 17 September 2017).

UNIVERSITY OF THE WITWATERSRAND. 2017b. ***Master of Science in Epidiomiology.*** Johannesburg: University of the Witwatersrand. [Online] available at https://www.wits.ac.za/publichealth/academic-programmes/postgraduate/master-of-science-in-epidemiology/ (Accessed 12 February 2018).

UNIVERSITY OF THE WITWATERSRAND LIBRARY*.* n.d. ***Scholarly Research and Related Resources.*** Johannesburg. University of the Witwatersrand Library. [Online] available at http://libguides.wits.ac.za/content.php?pid=284380&sid=2346960 (Accessed 10 March 2014).

UNSWORTH, J. 2006. ***Our Cultural Commonwealth: the report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences***. New York, NY: American Council of Learned Societies. [Online] available at http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf (Accessed 3 March 2015).

uPortal, 2009. Westminster, CO: Jasig. [Online] available at http://www.jasig.org/uportal (Accessed 6 February 2013).

USGS. n.d. ***USGS Data Management***. [Online] available at http://www.usgs.gov/datamanagement/describe/metadata.php (Accessed 19 August 2014).

USGS. 2013. ***Data Stewardship: roles and responsibilities, USGS***. Reston, VA: USG. [Online] available at http://www.usgs.gov/datamanagement/plan/stewardship.php (Accessed 18 August 2014).

VALIPOUR, M.H., AMIRZAFARI, B., MALEKI, K.N. & DANESHPOUR, N. 2009. A brief survey of software architecture concepts and service oriented architecture. In: ***Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2009, August 8-11, 2009, Beijing, China***, p. 34-38.

VAN DEN EYNDEN, V., ENSOM, T, WOLTON, A. & CORTI, L. 2014. ReShare and ReCollect: EPrints data repository solutions developed at the UK Data Archive. Presented at the ***OR2014 conference***, 13 June, Helsinki, Finland. [Online] available at https://core.ac.uk/display/39962556/tab/similar-list (Accessed on 25 September 2017).

VAN DER VAART, L. 2010. ***Collaboratories: connecting researchers: how to facilitate choice, design and uptake of online research collaborations.*** Utrecht, Netherlands: SURFfoundation. [Online] available at http://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2010/Collaboratories+Connecting+Researchers9april.pdf (Accessed 3 March 2015).

VAN DER WALT, A. 2017. ***Software Data Carpentry Instructor Training confirmations.*** E-mail received on 31 March 2017.

VAN DER WALT, I. 2017. ***Figshare-UP chat.*** E-mail received on 11 September 2017.

VAN DEVENTER, M. 2015. Research Data Management introduction. Presented at the ***Carnegie CPD 3 Workshop***, 21 November, University of Pretoria, Pretoria. [Unpublished].

VAN DEVENTER, M., PIENAAR, H., MORRIS, J. & NGCETE, Z. 2009. Virtual research environments: learning gained from a situation and needs analysis for malaria researchers. Presented at the ***African Digital Scholarship and Curation Conference***, 12-14 May, CSIR, Pretoria, South Africa. [Online] available at http://www.library.up.ac.za/digi/docs/mvdeventer_present.pdf (Accessed 17 February 2012).

VAN DEVENTER, M., BECKER, B., PIENAAR, H., MORRIS, J. & NYAKUNA, E. 2011. African researchers using natural products to improve lives. ***Library Intranet News,*** March, Issue 1. [Online] available at http://www.ais.up.ac.za/newsletter/libnewsmarch_11/African%20researchers.pdf (Accessed 27 September 2012).

VAN DEVENTER, M. & PIENAAR, H. 2009. Report on the 2nd African Digital Scholarship and Curation Conference. ***D-Lib Magazine***, July / August, 15(7/8). [Online] available at http://www.dlib.org/dlib/july09/vandeventer/07vandeventer.html (Accessed 4 January 2017).

VAN DEVENTER, M. & PIENAAR, H. 2015. Research data management in a developing country: a personal journey. ***International Journal of Digital Curation***, 10(2): 33-47. [Online] available at http://www.ijdc.net/index.php/ijdc/article/viewFile/10.2.33/406 (Accessed 4 February 2017).

VANNINI, P. 2008. Visual data. In: ***The Sage encyclopedia of qualitative research methods.*** Edited by Lisa M. Given. Thousand Oaks, CA: Sage Publications, Vol.2, p. 928-930.

VAN TILL, F. 2005. VRE Rapid Innovation: VRERI: kick-off and documentation***.*** Presented at the ***JISC Conference 2005***, 12 April, International Convention Centre, Birmingham, UK. [Online] available at http://bit.ly/MNW4eZ (Accessed 23 June 2012).

VAN WYK, B.J. 2005. ***Communities of Practice: an essential element in the knowledge management practices of an academic library as learning organisation.*** Masters thesis, University of Pretoria, Pretoria, South Africa.

VAN WYK, J. 2013a. ***Findings of an investigation into the essential research data that the University must manage: interviews with Deputy Deans Research of all the Faculties of the University of Pretoria: undertaken by the Department of Library Services from August-November 2013.*** Pretoria: Department of Library Services, University of Pretoria, 2013.

VAN WYK, J. 2013b. ***UP Policy for Research Data Management.*** E-mail sent to Prof Stephanie Burton on 14 March 2013.

VAN WYK, J. 2014a. ***Research Data Management Report.*** Compiled 31 July 2014***.*** Pretoria: Department of Library Services, University of Pretoria.

VAN WYK, J. 2014b. [First Draft] ***Research Data Management Policy.*** Pretoria: University of Pretoria.

VAN WYK, J. 2017. ***UP RDM policy.*** E-mail sent to Prof Stephanie Burton on 26 January 2017.

VAN WYK, J., KLEYN, L., & BUTLER-ADAM. 2017. [Final Draft] ***Research Data Management Policy.*** Pretoria: University of Pretoria.

VAN WYK, J. & VAN DER WALT, I. 2017. ***Criteria and evaluation of research data repository platforms at the University of Pretoria, South Africa.*** Presented at the ***eResearch Africa 2017 Conference***, 2-5 May, University of Cape Town, Cape Town. [Online] available at http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/Evaluation_repositories_UP_eRA2017_VanWyk_VanderWalt.pdf (Accessed 17 September 2017).

VINOGRADOV, S. & PASTSYAK, A. 2012. Evaluation of data anonymization tools. In: ***Proceedings of DBKDA 2012: The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications, held 29 February-5 March, 2012, Reunion Island.*** Wilmington, DE: International Academy, Research, and Industry Association (IARIA).

Virtual Research Environments: what is a VRE? 2011. [Wellington, NZ]: Victoria University of Wellington. [Online] available at http://ecs.victoria.ac.nz/EResearch/VirtualResearchEnvironments (Accessed 25 February 2012).

VOLLMAN, A.R., ANDERSON, E.T. & MCFARLANE, J. 2004. ***Canadian Community as partner.*** Philadelphia, PA: Lippincott Williams & Wilkins.

VOSS, A. & PROCTER, R. 2009. Virtual research environments in scholarly work and communications. ***Library Hi Tech***, 27(9): 174-190.

WASHINGTON UNIVERSITY IN ST. LOUIS. 2013. Scholarly communications. St. Louis, MO: Washington University in St. Louis. [Online] available at http://scholarlycommunications.wustl.edu/about/index.html (Accessed 12 February 2018).

WEAVER-HART, A. 1988. Framing an innocent concept and getting away with it. ***UCEA Review,*** 24(2): 11–12.

WhatsApp, 2012. Santa Clara County, CA: WhatsApp Inc. [Online] available at http://www.whatsapp.com/ (Accessed 28 September 2012).

What is SharePoint?, 2012. Redmond, WA: Microsoft Corporation. [Online] available at http://sharepoint.microsoft.com/en-us/product/capabilities/Pages/default.aspx (Accessed 19 October 2012).

WHITE HOUSE, OFFICE OF MANAGEMENT AND BUDGET. 1999. ***CIRCULAR A-110 REVISED 11/19/93 As Further Amended 9/30/99.*** Washington, DC: White House. [Online] available at http://www.whitehouse.gov/omb/circulars_a110#36 (Accessed 19 August 2014).

Which SURFshare are you looking for? n.d. [Online] available at http://www.surf.nl/en/oversurf/Pages/WhichSURFshareareyoulookingfor.aspx (Accessed 12 July 2012).

WHYTE, W.F., GREENWOOD, D.J. & LAZES, P. 1991. Participatory action research: through practice to science in social research. In: ***Participatory action research.*** Edited by William Foote Whyte. (Sage focus edition; Vol. 123). Newbury Park, CA: Sage Publications, p. 19-55.

WIGGINS, A., BONNEY, R., GRAHAM, E., HENDERSON, S., KELLING, S., LEBUHN, G., LITTAUER, R., LOTTS, K., MICHENER, W., NEWMAN, G., RUSSELL, E., STEVENSON, R. & WELTZIN, J. 2013. ***Data management guide for public participation in scientific research.*** Albuquerque, NM: DataONE.

WIKINDX, 2013. New York, NY: Sourceforge, Dice Holdings Inc. [Online] available at http://wikindx.sourceforge.net/index.html (Accessed 6 February 2013).

WILKINS-DIEHR, N. 2007. Special Issue: Science gateways: common community interfaces to grid resources. ***Concurrency and Computation: Practice and Experience***, 25 April, 19(6): 743-749.

WILKINS-DIEHR, N., GANNON, D., KLIMECK, G., OSTER, S. & PAMIDIGHANTAM, S. 2008. TeraGrid Science Gateways and their impact on science. ***Birck and NCN Publications****,* Paper 434. [Online] available at http://docs.lib.purdue.edu/nanopub/434 (Accessed 18 February 2015).

WILKINS-DIEHR, N., BARKER, M. & GESING, S. 2016. Science Gateways, Virtual Labs and Virtual Research Environments. Presented at ***eResearch Australia 2016***, 10-14 October, Melbourne, Australia. [Online] available at https://eresearchau.files.wordpress.com/2016/03/1430-nancy-wilkins-meyer-and-michelle-barker.pdf (Accessed 23 September 2017).

WILSON, A., RIMPILÄINEN, S., SKINNER, D., CASSIDY, C., CHRISTIE, D., COUTTS, N. & SINCLAIR, C. 2007. Using a Virtual Research Environment to support new models of collaborative and participative research in Scottish education. ***Technology, Pedagogy and Education***, 1 October, 16(3): 289-304.

WILSON, R.L. & ROSEN, P.A. 2003. Protecting data through ‘pertubation’ techniques: the impact on knowledge discovery in databases. ***Journal of Database Management***, April-June, 14(2): 14-26.

WIMPENNY, K. 2010. Participatory action research: an integrated approach towards practice development. In: ***New approaches to qualitative research: wisdom and uncertainty.*** Edited by Maggi Savin-Baden and Clarire Howell Major. Abingdon, Oxon, UK: Routledge, p. 89-99.

WINTER, R. 1989. ***Learning from experience: principles and practices in action research.*** Philadelphia, PA: Falmer Press.

WOLSKI, M. & RICHARDSON, J. 2011. A framework for university research data management. Presented at the ***CCA-EDUCAUSE Australasia Conference***, 3-6 April 2011, Sydney, Australia. [Online] available at http://hdl.handle.net/10072/39672 (Accessed 3 January 2017).

WOODSIDE, A.G. 2010. ***Case study research: theory, methods, practice.*** Wagon Lane, Bingley, UK: Emerald.

WOOLFREY, L. 2014. ***UCT research data management policy project: report.***
Cape Town: University of Cape Town. [Online] available at https://www.datafirst.uct.ac.za/images/docs/20140307-uct-rsearch-data-management-woolfrey.pdf
(Accessed 30 April 2014).

WorldCat, 2012. Dublin, OH: OCLC Online Computer Library Center Inc., 2012. [Online] available at http://www.worldcat.org/ (Accessed 20 October 2012).

WOUTERS, P. 1996. Cyberscience. ***Kennis en Methode***, 20(2): 155-186.

WOUTERS, P. & BEAULIEU, A. 2006. Imagining e-science beyond computation. In: ***New infrastructures for knowledge production: understanding e-science.*** Edited by C.M. Hine. Hershey, PA: Information Science Publishing, p. 48-70.

WRIGHT, C. 2016. ***National Integrated Cyberinfrastructure System (NICIS): an initiative in support of advancing research and innovation in the higher education sector: discussion document.*** [Sl.]: Universities South Africa. [Online] available at http://www.usaf.ac.za/wp-content/uploads/2016/11/Discussion-document-The-National-Integrated-Cyber-Infrastructure-System-CSIR-1.pdf (Accessed 23 September 2017).

WU, W., URAM, T., WILDE, M., HERELD, M. & PAPKA, M.E. 2010. Accelerating science gateway development with Web 2.0 and Swift. In: *TG '10*, ***Proceedings of the 2010 TeraGrid Conference held in Pittsburgh PA, 2-5 August 2010.*** New York, NY: ACM. [Online] available at http://www.mcs.anl.gov/uploads/cels/papers/P1765.pdf (Accessed 4 February 2012).

WULF, W.A., URAM, T., WILDE, M., HERELD, M. & PAPKA, M.E. 1989. The National Collaboratory: a White Paper. Appendix A. In: ***Towards a National Collaboratory: the unpublished report of an invitational workshop held at Rockefeller University, 17-18 March 1989.*** Chaired by Joshua Lederberg and Keith Uncapher. [New York: s.n.].

WUSTEMAN, J. 2009. Editorial: Virtual research environments: issues and opportunities for librarians. ***Library Hi Tech***, 27(2): 169-173.

XSED: EXTREME SCIENCE AND ENGINEERING DISCOVERY ENVIRONMENT. 2017a. ***XSEDE governance.*** [Sl.]: XSEDE. [Online] available at https://www.xsede.org/about/governance (Accessed 21 September 2017).

XSED: EXTREME SCIENCE AND ENGINEERING DISCOVERY ENVIRONMENT. 2017b. ***XSEDE resources.*** [Sl.]: XSEDE. [Online] available at https://www.xsede.org/ecosystem/resources (Accessed 21 September 2017).

XSED: EXTREME SCIENCE AND ENGINEERING DISCOVERY ENVIRONMENT. 2017c. ***Science Gateways.*** [Sl.]: XSEDE. [Online] available at https://www.xsede.org/ecosystem/science-gateways (Accessed 21 September 2017).

YANG, X. & ALLAN, R. 2006a. Realise e-research through Virtual Rresearch Environments. In ***Proceeding of the 4th International Conference on E-Activities***, Venice, Italy, 20-22 November, p.453-458. [Online] available at http://esc.dl.ac.uk/Sakai/papers/conf.pdf (Accessed 17 February 2012).

YANG, X. & ALLAN, R. 2006b. Web-Based Virtual Research Environments (VRE): Support Collaboration in e-Science. In: ***Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06,*** held 18-22 December 2006, Hong Kong Convention and Exhibition Centre, Hong Kong.

YANG, X. & ALLAN, R. 2007. Sakai VRE demonstrator project: realise e-research through Virtual Research Environments. ***WSEAS Transactions on Computers***, March, 2007, 6(3): 539-545. [Online] available at http://esc.dl.ac.uk/Sakai/papers/journal.pdf (Accessed 18 September 2013).

YANG, X. & ALLAN, R.J. 2010. Web-based Virtual Research Environments. In: ***Web-based Support Systems.*** Edited by JingTao Yao. London: Springer, p. 65-79.

YAO, Y.Y. 2003. A framework for Web-based Research Support Systems. In: ***Proceedings of the 27th Annual International Computer Software and Applications Conference (COMPSAC'03).*** [Los Alamitos, CA]: IEEE Computer Society, p. 601-606.

YAO, Y.Y. 2004. Web-based research support systems. Presented at the ***Second International Workshop on Web-based Support Systems*** in conjunction with IEEE/WIC/ACM WI/IAT'04, 20 September 2004, Beijing, China.

YIN, R.K. 2003. ***Case study research: design and methods.*** (3rd ed, Vol.5), Thousand Oaks, CA: SAGE.

YIN, R.K. 2009. **Case study research: design and methods.** 4th Edition. (Applied Social Research Methods Series, v.5). Thousand Oaks, CA: SAGE.

YIN, R. 2012. ***Applications of case study research.*** Thousand Oaks, CA: SAGE.

YouTube, 2012. San Bruno, CA: YouTube, LLC. [Online] available at http://www.youtube.com/ (Accessed 20 October 2012).

ZEENOV Agora, 2013. Somerville, NJ: Zeenov Inc. [Online] available at https://www.zeenov.com/ (Accessed 22 March 2013).

ZHONG, N., LIU, J.A. & YAO, Y. 2003. Web Intelligence (WI): a new paradigm for developing the Wisdom Web and Social Network Intelligence. In***: Web Intelligence.*** Edited by Ning Zhong, Jiming Liu and Yiyu Yao. Berlin, Germany: Springer, p.1-16.

ZIMMER, N. 2017. Figshare engagement across South Africa. ***NeDICC Workshop,*** 13 September, CSIR, Pretoria, facilitated online. [Unpublished].

ZINS, C. 2007. Conceptual approaches for defining data, information, and knowledge. ***Journal of the American Society for Information Science and Technology***, 15 February, 58(4): 479-493.

Zotero, n.d. Fairfax, Virginia: Roy Rosenzweig Center for History and New Media, Department of History and Art History, George Mason University. [Online] available at http://www.zotero.org/ (Accessed 9 August 2012).

ADDENDUM A: TERMINOLOGY

Terms / Concepts	Description
Access Grid	*"An ensemble of resources including multimedia large-format displays, presentation and interactive environments, and interfaces to Grid middleware and to visualization environments. These resources are used to support group-to-group interactions across the Grid"* (Universidad de Cordoba, n.d.).
application	An application is a software programme "designed to perform a specific function directly for the user, or in some cases for another applications program", e.g. word processing software, database programmes, e-mail programmes, games, Web browsers, development tools, drawing programmes, imaging programmes and communication programmes (Rose, 2007). "The word 'application' is used because each program has a specific application for the user" (Christensson, 2008).
Application Programming Interface (API)	"An application program interface (API) is code that allows two software programs to communicate with each other. The API defines the correct way for a developer to write a program that requests services from an operating system (OS) or other application" (Essential guide to API management and application integration, 2017).
Collaborative Virtual Environment	*"A multi-party virtual environment which allow a number of users to share a common virtual space, where they may interact with each other and the environment itself"* (Goebbels & Lalioti, 2001: 155).
collaboratory	*"An organizational entity that spans distance, support rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artefacts, and tools required to accomplish research tasks"* (Bos et al., 2007: 653, 656).
cyberinfrastructure	The term *"refers to an infrastructure of distributed computer, information and communication technologies"* (Atkins et al., 2003: 5).
cyberscience	*"All scholarly and scientific research activities in the virtual space generated by the networked computers and by advanced information and communication technologies in general"* (Nentwich, 2003: 22).

cloud computing	“Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS)” (Rouse & Bigelow, 2017).
Communities of Practice	*“*A network of people emerging spontaneously, and held together by informal relationships and common purpose, that share common knowledge or a specific domain, expertise and tools, and learn from one another*”* (Van Wyk, 2005: 7).
data archiving	Data archiving is the process of retention and storage of valuable data for long-term preservation, so that the data will be protected from risk (i.e. loss, or corruption), and will be accessible for future use (CODATA Workshop on Archiving Scientific & Technical (S&T) DATA, 20-21 May 2002, Pretoria, South Africa: report, 2002; Müller, 2009; Rouse, 2010).
data curation	Data curation is the active and ongoing activities that data stewards engage in to add value to research data throughout its entire lifecycle so that the data are meaningful and useful to scholarship, research and education, and available for discovery and re-use (Choudhury, 2013; Cragin et al., 2007; DCC, 2014c; Dempsey, 2007; Lord and MacDonald, 2003: 12; UCLA Library, 2014; UC San Diego, 2014; University of Minnesota Libraries, 2014).
data governance	Data governance can be described as a strategic function and is concerned with the people managing the data, which includes goals, policies, shared decision making, planning, strategies, and processes followed (Haines, 2012; Rouse, 2007; DAMA Dictionary of Data Management, 2011).
data management	“Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets” (DAMA International, 2007).
Data Seal of Approval	The Data Seal of Approval is an international certification assigned to data repositories for the safeguarding of data and “to ensure high quality”, and give guidelines on “reliable management of data for the future”, without necessitating the application “of new standards, regulations or high costs” (Data Seal of Approval, n.d.).
data stewardship	Data stewardship can be described as a specific approach to data management; it is about taking responsibility for

	data sets, and is a tactical function, that is executed against specific data criteria (Haines, 2012; USGS, 2013).
demonstrator	A working prototype (Allan, 2009: 167)
digital repository	“In simplest terms”, a digital repository is a database / software “where digital content” and assets can be stored, searched and retrieved for later use (Hayes, 2005).
e-collaboration	e-collaboration is “collaboration among individuals engaged in a common task using electronic technologies” (Kock et al., 2001: 1)
e-learning	*“*e-Learning is the use of technology to enable people to learn anytime and anywhere*”* (Commissionerate of Collegiate Education, 2017).
e-learning system	*“*A comprehensive software package that supports courses that depend on the” World Wide Web “for some combination of delivery, testing, simulation, discussion, or other significant aspect*”* (Robson, 1999: 271).
e-Research	E-Research can be defined as a broad term that extends e-Science. It is a form of scholarship conducted in a networked environment that includes all ICTs that support researchers in their research process. This includes all forms of non-computational e-Science, consisting of a wide variety of new technologies, tools and computer networks, which can be used collaboratively by researchers and that can be co-located or separated by distance globally.
e-Science	e-Science can be described as “global collaboration in key areas of science (Hey and Trefethen, 2003: 1017; Jankowski, 2007: 551), “the next generation of scientific problems, and the collaborative tools and technologies that will be required to solve them” (Hey and Trefethen, 2008: 15). E-Science according to (Beaulieu and Wouters, 2009: 55, 56) includes “the sharing of computational resources, [high performance computing], distributed access to massive datasets, and the use of digital platforms for collaboration and communication”, but does not cover non-computational e-Science and research not reliant on high performance computing (HPC).
experimental workflow	See Scientific workflow
framework	“A software framework helps facilitate software development by providing generic capabilities that can be changed or configured to create a specific software application” (Jamison, Bortlik and Hanley, 2013).

Grid	“Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” *(*Foster, Kesselman & Tuecke, 2001: 200).
Grid Computing	“Grid computing is a form of distributed computing in which use is made of a ‘grid’ composed of networked, loosely-coupled computers, data storage systems, instruments etc” (Allan, 2009: 133).
groupware	“Groupware is software specially written to be used by a group of people connected to a network and help them carry out a particular task; it provides useful functions such as a diary or electronic mail that can be accessed by all users” (Collin, 2002: 224-225).
Instant Messaging	Instant messaging is the sending of a message “from one person to another that appears immediately on the recipient’s computer, allowing a text communication” (Levine, Young and Baroudi, 2005: 363).
interface	“It is the way a user interacts” with software (e.g. an application, programme or a website) or hardware” (Christensson, 2009; Rouse, Churchville and Dang, 2016)
middleware	“Middleware offers general services that support distributed execution of applications. The term middleware suggests that it is software positioned between the operating system and the application. Viewed abstractly, middleware can be envisaged as a "tablecloth" that spreads itself over a heterogeneous network, concealing the complexity of the underlying technology from the application being run on it” (Puder, Römer and Pilhofer, 2006: 21).
Open Archives Initiative Protocol for Metadata Harvesting	“This protocol provides the basis for an information discovery environment that relies on transferring metadata en masse from one server to another in a network of information systems” (Marshall, 2002: 24).
Open Grid Computing Environment (OGCE)	The OGCE is National Science Foundation-funded collaboration comprised of Indiana University, San Diego State University, San Diego Supercomputer Center, and the Texas Advanced Computing Center “that spans diverse research and development efforts”, for example “portlet development environments; Grid computing abstraction layers; advanced Grid services for information, science application, and data management; and services for group collaboration"(Alameda et al., 2007: 940). These elements are then integrated using “standards such as JSR 168-compatible portlets and Web Services in a common portal architecture” (Alameda et al., 2007: 940)

platform	A platform includes any hardware or software upon which software applications or services can be built and run (Bigelow and Rouse, 2016; Jamison, Bortlik and Hanley, 2013; Martin, 2014).
portal	“An integrated and personalized web-based interface to information, applications and collaborative services” (Chochan, 2005).
research data lifecycle	A cycle that illustrates the flow of data during the research process through a number of components (stages), for example: creating data, analysing data, preserving data, giving access to data, and re-using data (Ball, 2012; Beagrie, 2004; DCC, 2014b; Wiggins et al., 2013: 1-14). These components (stages) can take place in any number of different sequences, with some occurring simultaneously and some repeated more than once (Wiggins et al., 2013: 2).
research lifecycle	A model that represents the life course of a larger system, such as the research process, through a series of sequentially related stages or phases in which information is produced or manipulated (Humphrey, 2006).
Research Data Management	Research Data Management is the process of controlling and organising the data generated during a research project, and covers the entire data lifecycle, which includes the planning of the investigation, conducting the investigation, storage and backing up of the data as it is created, preserving the data long-term, after the research investigation has concluded, and making the data accessible for future use (Penn State University Libraries, 2014; Texas A & M University Libraries, n.d.; University of Tennessee Libraries, 2014).
scholarly communication	“Scholarly Communication pertains to the creation, transformation, dissemination and preservation of knowledge”, encompassing “teaching (promotion and transmission of knowledge), “research (creation of new knowledge) and scholarly activities” (Washington University in St. Louis, 2013; Bernard Becker Medical Library, 2017).
Science Gateway	A Science Gateway can be described as a community-developed bundle of tools, applications, and data collections customised to meet the needs of a targeted community, which are integrated via a portal, or a collection of applications (Indiana University, 2012; Wilkins-Diehr, 2007: 743; Yang and Allan, 2010: 69). A science gateway provides an interface between a researcher (or

	community) and distributed computing infrastructures (LPDS, 2013).
scientific workflow	“Amalgamation of scientific problem-solving and traditional workflow techniques” (Singh and Vouk, 1996); “A blanket term to describe series of structured activities and computations that arise in scientific problem-solving” (Singh and Vouk, 1996); “A scientific workflow is the process of combining data and processes into a configurable structured set of steps that implement semi-automated computational solutions of a scientific problem” (Altintas et al., 2006: 468).
Service-Orientated Architecture	“Service-Orientated Architecture is an approach to joining up independent services to provide integrated capabilities. A key aspect of the architecture is to maximise the re-use of common services and middleware, including portlets”, with “web service interfaces for all these components” (Allan, 2009: 16-17).
South African Research Information Services Project (SARIS)	A South African project “started inter alia because of the extremely high costs to South African research institutes and university libraries to access the global research literature”, and tasked with the specific aim of designing a possible structure for an eResearch Service for South Africa (Bothma, Pienaar and Hammes, 2008: 272).
Virtual Learning Environment	The concept Virtual Learning Environment is synonymous with an E-Learning system. A virtual learning environment (VLE) is a designed information and social space, which is explicitly represented (from text-based interfaces to the most complex 3D graphical output, where students are not only active but are also actors. VLEs are not restricted to distance education but also integrate multiple tools and overlaps with the physical environment (Dillenbourg, 2000: 3-12).
Virtual Organisation	A group of people (or institution) that have authorised access to sets of resources, and which coordinates the sharing of these resources, and coordinates problem solving (Allan, 2009: 82); Foster, Kesselman and Tuecke (2001).
Virtual Research Community	“A Virtual Research Community is a group of researchers, possibly widely dispersed, working together and facilitated by a set of online tools, systems and processes interoperating to support collaborative research within or across institutional boundaries” (Pothen, 2004: 22).
Virtual Research Environment	A Virtual Research Environment (VRE) consists of a common, flexible, technological and collaborative

	framework into which online tools (or applications), technologies, services, data, and information resources (e.g. articles, concept papers, drafts etc.) interoperating with each other, can be plugged, to enable collaboration and to support and enhance large and small scale processes of research, which are performed by researchers in multidisciplinary contexts and across organisational and geographical boundaries.
Web 2.0	"Refers to a supposed second-generation of Internet-based services - such as social networking sites, blogs, wikis, communication tools, and folksonomies - that let people collaborate and share information online in ways previously unavailable." (The Hatchergroup, 2008). Seven principles describe Web 2.0: Web as platform, harnessing collective intelligence, data is the next 'Intel' inside, end of the software release cycle, lightweight programming models, software above the level of single device, and rich user experiences. (O'Reilly, 2005)
Web-based research support systems	Web-based research support systems (WRSS) are systems that are done via the web. WRSS aims to develop "new and effective tools for research institutions, researchers and scientists" so as to support their research activities and assist them in the improvement of their research quality and productivity (Tang et al., 2003: 21).

ADDENDUM B: DATA MANAGEMENT PLAN

Data Management Categories	
Administrative Data	
Funder	NRF
Project Name	The Relationship between Research Data Management and Virtual Research Environments
Project Description	A PhD degree study at the University of Pretoria consisting of an investigation of the place of Research Data Management (RDM) within a Virtual Research Environment (VRE), through a literature study, testing and prototyping, and interviews.
Principal Investigator / Researcher	B.J. van Wyk, Assistant Director Research Data Management, University of Pretoria
Principal Investigator / Researcher ID	0000-0003-2869-4377 (ORCID)
Project Data Contact	See PI / Researcher
Date of First Version	15/08/2015
Date of Last Update	07/10/2017
Related Policies	None
Data Collection	
What data will you collect or create?	• Data will be qualitative in nature. • The project will not be using existing data.
How wil the data be collected or created?	• Data will be collected via face-to-face and video interviews, as well as through meeting notes and e-mail correspondence. • The interview will consist of 49 interview questions, of which 23 questions are directed to student researchers, 9 directed to VRE Managers, 12 directed to the VRE designer and 5 directed to the librarian / information specialist. The questions to the student researchers will be divided into questions on the VRE, questions on RDM and general questions on VREs. • The target population and sample will be identified from two case studies at the University of Pretoria. • Responses to the questions will be captured through audio-recordings and transcriptions on MS Word. • File formats: Sound File (.mp3) and MS Word for text documents (.docx), and image files (.png). • Data volume: approximately 400 MB. Storage space is not an anticipated problem.
Documentation and Metadata	
What documentation and metadata will accompany the data?	• A project information sheet, and methodology description need to accompany the data. This will be captured in MS Word format. • Metadata will be created. Dublin Core will be used as metadata standard.

Ethics and Legal Compliance	
How will you manage any ethical issues?	• This project received ethical clearance from the University of Pretoria. • Informed consent will be gained from participants prior to data collection. • Confidentiality: The project will be clearly explained to each participant and personal details will only be captured for administrative purposes and will not be disclosed. • Personal details or information revealed via responses to open-ended questions will be anonymised and de-identified. • Respondents will be asked to sign a consent form, which specifies what data will be collected and how it will be managed and used.
Storage and Backup	
How will the data be stored and backed-up during the research?	• Storage space is not an anticipated problem. • Data will be stored on the University of Pretoria's institutional instance of Google Drive, on this researcher's personal computer's hard drive, as well as on an external hard drive.
How will you manage access and security?	• A user ID/password is required to access the University of Pretoria's instance of Google Drive. • A user ID/password is required to access backed-up data in all other devices. • Security of sensitive/personal data: this is not really an anticipated problem; nevertheless, data will be anonymised and de-identified should it be necessary.
Selection and Preservation	
Which data should be retained, shared and/or preserved?	• The anonymised interview data will be kept for 10 years on the University of Pretoria's Google Drive, with a backup on an external hard drive.
Any restrictions on data sharing required	• Anonymised processed data will be shared freely and openly, through the University of Pretoria's institutional repository (UPSpace). Anonymised raw data will be made available on request through e-mail, and on an institutional data repository as soon as this has been made available.
Responsibilities and Resources	
Who will be responsible for data management?	• The principal investigator (PI)
What resources will you require to deliver your plan?	• No additional resources will be required.

ADDENDUM C: DATA DOCUMENTATION SHEET

Project background

Virtual Research Environments (VREs) as technology frameworks to facilitate collaborative research projects have been used by a number of universities and research institutions globally. Technologies used to build VREs have tended to vary significantly, resulting in fragmentation and interoperability, which necessitates agreed standard platforms and configurable modules (Voss and Procter, 2009: 176). There is thus a need for the formalisation of a conceptual model of a VRE that can be used repeatedly in different contexts and different subject fields. Furthermore, research data is recognized internationally as a vital resource, which needs to be preserved for future research. VREs offer the ideal instruments that could be used in the management of research data. The aim of the study was to compile a conceptual model of a VRE that indicates the relationship between VREs and Research Data Management (RDM) – an essential component of a VRE.

Topic

The relationship between between Research Data Management and Virtual Research Environments

This study was a topic of a PhD thesis in the Department of Information Science at the University of Pretoria.

Central Research Question

The central research question was: How can a Virtual Research Environment be conceptualised to indicate the role of Research Data Management (RDM) within a VRE?

To answer this question a number of sub-questions were asked. These are:

- What is a VRE?
- What is the current state of VRE research in the world?
- What are the generic components that make up a VRE?
- How does a VRE support a research cycle?

- What is Research Data Management?
- Why should a VRE be an essential technological and collaborative framework for the management of research data?
- To what extent can the components identified be formalised into a conceptual framework
- Where would RDM as component be placed?
- To what extent can this model be generalised for use in other environments?

Investigator

Barend Johannes van Wyk
Contact: johann.vanwyk@up.ac.za

Population studied

Two case studies from the University of Pretoria were identified.
Respondents from Case Study A included: 1 VRE Manager, 1 VRE Champion, and 5 postgraduate researcher students.
Respondents from Case Study B included: 1 VRE Manager, 4 postgraduate students, and 1 librarian.
The VRE designer of both case studies was also a respondent.

Data collection

This study consisted of interviews, testing and prototyping and a literature review.

Sampling

Purposive sampling was chosen to identify respondents, because of the researcher's knowledge of the researchers involved, as well as their roles and characteristics within the VREs. Through this process two case studies at the University of Pretoria were chosen.

Instruments and software used

The study made use of an interview questionnaire, which was created by the researcher of this study. These interviews were then audio recorded (in mp3 format) and transcribed in MS Word (docx format). Notes were taken during meetings with members of the group, and e-mail correspondence also rendered valuable information.

Timespan of the interviews

Each interview lasted for approximately 1 hour

Data files

Data for this study comprise 15 raw data files (approximately 400 MB) and the processed data are included in the thesis.

Data validation

Transcriptions of the individual interviews were sent to the individual interviewees for clarification, validation and commentary, and data was corrected and adjusted in line with recommendations from these respondents.

Data confidentiality, access and use

Raw anonymised data will be placed in an institutional data repository or archive after publication. It will be freely available for re-use and sharing.

Dataset details

The datasets are in MS Word format, with the file type in .docx format.

METADATA

Abstract

This study focuses on Virtual Research Environments (VREs) as ideal technological and collaborative frameworks for the management of research data. The aim of the study was to compile a conceptual model of a VRE that indicates the relationship between VREs and Research Data Management (RDM) – an essential component of a VRE.

In the first part of the study, a literature review was conducted by focusing on four themes: VREs and other concepts related to VREs; VRE components and tools; RDM; and the relationship between VREs and RDM. The first theme included a discussion of definitions of concepts, approaches to VREs, their development, aims, characteristics, similarities and differences of concepts, an overview of the e-Research approaches followed in this study, as well as an overview of concepts used in this study. The second theme consisted of an overview of developments of VREs in four countries (United Kingdom, USA, The Netherlands, and Germany), an indication of the differences and similarities of these programmes, and a discussion on the concept of research lifecycles, as well as VRE components. These components were then matched with possible tools, as well as to research lifecycle stages, which led to the development of a first conceptual VRE framework. The third theme included an overview of the definitions of the concepts 'data' and 'research data', as well as RDM and related concepts, an investigation of international developments with regards to RDM, an overview of the differences and similarities of approaches followed internationally, and a discussion of RDM developments in South Africa. This was followed by a discussion of the concept 'research data lifecycles', their various stages, corresponding processes and the roles various stakeholders can play in each stage. The fourth theme consisted of a discussion of the relationship between research lifecycles and research data lifecycles, a discussion on the role of RDM as a component within a VRE, the management of research data by means of a VRE, as well as the presentation of a possible conceptual model for the management of research data by means of a VRE. This literature review was conducted as a background and basis for this study.

In the second part of the study, the research methodology was outlined. The chosen methodology entailed a non-empirical part consisting of a literature study, and an empirical part consisting of two case studies from a South African University. The two case studies were specifically chosen because each used different methods in conducting research. The one case study used natural science oriented data and laboratory/experimental methods, and the other, human orientated data and survey instruments. The proposed conceptual model derived from the literature study was

assessed through these case studies and feedback received was used to modify and/or enhance the conceptual model.

The contribution of this study lies primarily in the presentation of a conceptual VRE model with distinct component layers and generic components, which can be used as technological and collaborative frameworks for the successful management of research data.

Keywords

Virtual Research Environments, Research Data Management, core components, RDM components, pluggable components, research lifecycle, research data lifecycle, conceptual model.

Collection period

Data from the interviews were collected over a five-month period, starting in September 2015 and ending in January 2016.

Location

University of Pretoria, Pretoria, South Africa.