Research Data Management (RDM) in agricultural research institutes: a literature review

Emily Ng'eno[1]
Moi University, Kenya
ngenoemily28@gmail.com

and
Stephen Mutula[2]
University of KwaZulu-Natal
mutulas@ukzn.ac.za

Abstract

This article presents a survey of literature on Research Data Management (RDM) with focus on agricultural research institutes. This is to help the understanding of core issues in RDM such as legal, policy and regulations; skills set and infrastructure in order to strategically position the agricultural sector in the knowledge economy and also help in reducing duplication of effort, promoting innovation, minimizing loss or destruction of research data sets and that ensuring compliance with funders' requirements. The author argues that while RDM has been widely embraced in developed countries Africa is lacking behind. The literature reviewed in this article seem to suggest that legal, policy and regulatory framework in agricultural research sectors are either nonexistent or outdated. This is exacerbated by inadequate technical infrastructure, human resources capacity, and paucity of national or international partnerships. As a result, research data sets within agricultural institutes are poorly managed. The establishment of a legislative and policy framework for RDM; capacity building programs, and improvement of technical infrastructure are highly recommended.

Keywords: Research Data Management, data curation, research data, agricultural research institutes, Kenya

1. Introduction

The purpose of this literature review is to examine Research Data Management (RDM) with a view to understanding legal, policy and regulations in RDM. Research data are valuable resources that need to be properly managed by research institutes to contribute to technology innovation and transfer in the agricultural sector. Besides research data being intricate and complex, they are irreplaceable, expensive and time-consuming to replicate. In addition, the research institutes must exercise diligence and ensure data accuracy and precision in the collection, description, preservation, access, reuse and sharing of research data (Fellous-Sigrist 2015; University of California, Los Angeles n.d). The Government of Canada (2016) maintains that the ability to preserve, access, reuse and build upon research data is critical to the advancement of science and scholarship. In addition, research data is instrumental in supporting innovative solutions to economic and social challenges, and holds tremendous potential for productivity, competitiveness, and quality of life. Therefore, there is a need for research institutes to engage in RDM cannot be underestimated. Ray (2014) defines Research Data Management

1. Emily Ng'eno PhD is a lecturer in the Department of Library, Records Management and Information Studies, Moi University , Kenya. She graduated with a PhD from the University of KwaZulu-Natal in 2018.
2. Stephen Mutula PhD is Professor of Information Studies and Dean of Humanities, University of KwaZulu-Natal, South Africa .He supervised Emily's PhD.

(RDM) as the collection, organization, validation, and preservation of data for analysis, discovery, sharing, reuse, and transformation. Fundamentally, RDM consists of different activities and processes associated with data creation, storage, security, preservation, retrieval, reuse and sharing taking into account technical capabilities, ethical considerations, legal issues, human resource capability and government frameworks.

The literature reviewed in this article on RDM is informed by Community Capability Model Framework (CCM Framework) and Data Curation Centre (DCC) Lifecycle Model (Ng'eno 2018:22). Community Capability Model (CCM) Framework was developed by the United Kingdom Office for Library and Information Networking (UKOLN), University of Bath and Microsoft Research to assist research funders, institutions and researchers in growing the capability of their communities to perform data-intensive research (Lyon, Ball, Duke & Day 2012). It focuses on eight capability factors that include collaboration, skills and training, openness, technical infrastructure, common practices, economic and business, legal and ethical, and academic issues representing human, technical and environmental issues. Key variables in CCM framework relevant to this literature review on RDM include: openness, skill and training, technical infrastructure, legal and policy issues, and collaborative partnerships.

The DCC Lifecycle Model on the other hand promotes a lifecycle approach to the management of digital materials to enable their successful curation and preservation from their initial conceptualization to either disposal or selection for re-use and long-term preservation (Higgins 2008). According to Data Curation Centre (DCC) Lifecycle Model, data curation includes the data capture, appraisal, description, preservation, access, re-use and transformation of research data. Higgins (2008) emphasizes that Data Curation Centre (DCC) Lifecycle Model advocates for maintenance of authenticity, reliability, integrity and usability of digital material which in return ensures quality of RDM. The review is underpinned by the CCM Framework and DCC Lifecycle Model because the former addresses the community/institutional capabilities which incorporate skill and training, technical infrastructure, legal and policy issues, collaborative partnerships and openness which contribute a lot to RDM practices and the later equally prescribes the activities to be done in RDM such as data curation involving capture, appraisal, preservation, access and re-use. These activities are the core functionalities in RDM.

Therefore, it can be noted that RDM it is imperative in agricultural research institutes for the reason that if research data is managed well, it will promote access, sharing and reuse.

2. Research Data Management (RDM)

The availability of huge amounts of new data, often referred to as a data deluge, impacts the method in which research is carried out for societal benefit. RDM brings benefits to researchers and research institutes in many ways (Lewis 2010; and Dora and Kumar 2015) such as:

- ***i.*** Ability to share research data, minimizing the need to repeat work in the field or laboratory;
- ***ii.*** Research data gathered at considerable cost is not lost or inadvertently destroyed;
- ***iii.*** Retrieval, comparison, and co-analysis of data from multiple sources can lead to powerful insights;
- ***iv.*** New research themes can emerge from re-analysis of existing data or comparisons with new data;
- ***v.*** Long-term preservation of data provides for validation check of the data and this enhances the credibility and transparency of the research data used;
- ***vi.*** By opening research data sets for the public, there is visibility of the host institution and its researchers;
- ***vii.*** Research funders are increasingly requiring researchers to deposit their research data for proper curation, full utilization, preservation, and reuse (Heidorn 2011; Ingram 2016; Lyon, Patel, and Takeda 2014).

Furthermore, Ray (2014) explains that the sharing of research data increases the return on large investments, advances human knowledge, promotes economic development and reduces costly data duplication. Open access has emerged as one way of sharing research data to promote the advancement of knowledge and technology transfer. Organization for Economic Co-operation and Development (OECD) (2007) points out that the principles and guidelines of open access include openness, flexibility, transparency and legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability; and sustainability. RDM has become even more widely endorsed in agricultural research institutes and increasingly supported by the mandates of research funders who are keen to see the greatest possible returns on investments both in terms of quality of research output and the reuse of research data.

2.1 Legal, policy and regulations in RDM

In essence, RDM legal, policy and regulation framework should respond to a number of RDM drivers such as: storage, security, preservation, quality, compliance, sharing, and jurisdiction in order to enhance management, sharing and reuse of research output (Pinfield, Cox and Smith, 2014; Higman and Pinfield 2015).

Smith (2014) in a study on data governance argues that legal environment surrounding research data lags behind hindering the ability to develop best practices for data management, sharing and use. Moreover, RDM legal environment includes laws, regulations, and policies associated with data, as well as strategies for data quality control and management in the context of agricultural research institutes. In this respect, RDM legal framework ensures that data can be trusted through facilitation of RDM governance by adoption of appropriate technical standards, practices and architecture that will necessitate management, sharing and reuse of research data. A study by Fitzgerald, Pappalardo and Austin (2008) on RDM legal and policy guide advances the view that RDM legal framework should be comprehensive, coherent and precise. In addition, Karick (2014) points out that RDM framework should clearly state ownership and rights associated with research data in order to minimize disputes and provide accountability for research data.

Policy and regulations affecting RDM must be developed to address data curation, quality and security as well as ethical requirements, human resource capacity, technical infrastructure and collaborative partnerships at every stage of RDM. This should result in the development and operationalisation of an effective and efficient collection, appraisal, preservation, access, reuse and sharing of research data (Cox and Pinfield 2014). An RDM study conducted through an online survey of 145 research funders, national bodies and agricultural research institutes by Mossink, Bijsterbosch, and Nortier (2013) focusing on Support Infrastructure Models for RDM (SIM4RDM), underscored the fact that a number of agricultural research institutes with RDM policy is growing, however there is a need for institutional policies on RDM to have the following elements:

a) Responsibilities and roles;
b) Access and reuse of data;
c) Long term preservation/curation;
d) Security;
e) Open accessibility and availability of data;
f) Protection of legitimate subjects of research data (embracing informed consent, anonymity and confidentially).
g) Provision of mechanisms for storage, back up, registration;
h) Training, support, advice;
i) Copyright and intellectual property issues (copyright, patents, trademarks and design rights);

j) Embargo period for open accessibility;
k) Methods used to share guidelines/restrictions on data;
l) Destruction of records;
m) Removal or transport of data;
n) Preferred Licenses for data.

Similarly, the study concluded that there is a need for policies and guidelines to govern ownership of data created within research institutes and also raise awareness amongst researchers regarding important research data and related aspects as this was a problematic area (Mossink, Bijsterbosch, and Nortier 2013). In addition to the elements listed by Mossink, Bijsterbosch, and Nortier (2013) on what the policy should entail, Smith (2014) posits that RDM policy and regulation framework should address issues such as copyright, data licensing, data security, data privacy, and mind-set (researchers to accept the idea of their data being made available for reuse). Therefore, policy development is a cross-institutional process and by initiating the conversation about RDM policy, it should be an all-inclusive with the goal to facilitate effective and efficient management, sharing and reuse of research output (Erway 2013).

3.2 Data curation

Data curation as one of the key aspects of RDM involving research data capture, appraisal, description, preservation, access, reuse and transformation. Data curation is the active and on-going management of data throughout its lifecycle of interest and usefulness to research to enable data discovery and retrieval, maintain quality, add value and provide for reuse over time (Palmer, Cragin, Heidorn, and Smith 2007). Whyte and Allard (2014) acknowledged the challenges in data curation with regard to lack of legal framework, standards or procedures to reference and define mandatory guidelines when curating data.

3.2.1 Data capture

Australian National Data Service (n.d) defines data capture as the process of collecting data which will be processed and used later to fulfill certain purposes. Research data with good metadata attached at the point of capture can expedite data sharing, publishing and citation. Metadata capture is of great value simply because the more information there is about data, the greater the value of the data whether automatic or manual. Consequently, Australian National Data Service (n.d) buttress that data capture tools should:

a) Provide processes of organizing and structuring data files;
b) Have data validation components to ensure that captured data meet required types and ranges;
c) Enable open and flexible formats where good conversion tools exist; and
d) Allow data to be moved to its destination efficiently and with high quality.

Higgins (2012) advances the view that relationships developed between the researcher and the information professional is very crucial in planning research data capture, since the former is concerned with capturing data, while the latter is concern with receiving and ensuring data capture.

3.2.2 Data appraisal

Appraisal and selection is the process of evaluating research data in order to decide which to retain over the long term, which to retain for the meantime and which to discard (Higgins 2012; International Standards Organisation (ISO) 2001). Appraisal and selection policy need to ensure consistent, transparent and accountable decision making. Whyte and Wilson (2010) state that appraisal and selection policy must fit legal requirements, for example, relating to privacy and intellectual property rights, Public Records Acts, national data policies and codes of conducts adopted by the host institution or agricultural research institutes or funders. Niu (2014) and Eaker (2016) came up with a framework that synthesized traditional archival appraisal methodologies

and elements identified from existing appraisal policies with an intention to serve as high-level guidance for individual institutions to create their local appraisal/selection policies for research data. Niu (2014) and Eaker (2016) appraisal/selection criteria include:

a) Mission alignment: whether the resource supports the mission and falls within the scope of the collection policy of a preserver or the institution;

b) Value of digital resources: can either be primary or secondary value as Tibbo (2003); Schellenberg (1956); and UK National Archives (2012) point out; that primary values include administrative, fiscal, and legal value while secondary values includes evidential and information values.

c) Cost: It could be costs in acquiring, housing, preserving, and processing the collection to make it accessible also assess whether value of the data exceeds costs;

d) Feasibility: Feasibility of preservation is often determined by the technical capacity which include metadata and documentation, file formats and protection mechanisms such as password, digital signatures or encryption (National Archives and Records Administration (NARA) 2007).

Whyte and Wilson (2010) observed that in agricultural research institutes, research data librarian and archivist should assume the role and responsibility of setting appraisal/selection criteria and appraisal policy in consultation with stakeholders, especially researchers and local data managers in order to know how data would be assessed and how they would increase their enduring impact.

3.2.3 Data description

Research data description is the documentation that accompanies the research data which makes it discoverable and usable over time thus metadata standards exist to provide standardized descriptions, for example, Dublin Core, and computer software (The University of Western Australia 2016). Proper description and documentation of research data allows users to understand and track important details, in addition to describing research data using metadata facilitates, searching and retrieval in data repositories. Regarding the contents of metadata in describing research data, Cornell University Library (n.d) gives an example of the content of metadata such as contact information, geographic locations, units of measure, abbreviations or codes used in the dataset, instrument and protocol information, survey tool details and much more. Such detailed metadata content facilitates data curation which results in quality RDM. Data description is imperative in agricultural research institutes in terms of fully describing research data for easy accessibility, searching as well as retrieval and preservation.

3.2.4 Data preservation

Data preservation means securing permanent access to the original research data from the finished research project and general characteristics of data preservation is data accessibility to others for verification or for sharing or collaboration within the scientific community (Kruse and Thestrup 2014). Subsequently, it is imperative that long-term preservation and protection of sensitive data are vital characteristics of data preservation actions in agricultural research institutes. Furthermore, policies and requirements regarding where to deposit research data and the retention period should be clearly stipulated as well as address the gap between short-term access and long-term preservation with reference to the type of research data (RECODE Project Consortium 2014). Activities that support the preservation process should be planned to include administrative procedures required before undertaking preservation activities and the technical requirements of preservation.

A survey carried out by Mossink, Bijsterbosch and Nortier (2013) on Support Infrastructure Models for RDM (SIM4RDM) in Europe revealed that institutional repositories were also deployed to store finished datasets, for example, the Netherlands, the UK and Finland have well-

established data archived for storage. Research data preservation is an ongoing process that should be planned throughout its lifecycle to include the following activities: validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures (Data Curation Centre 2004). These preservation actions should ensure that research data remain authentic and reliable while maintaining their integrity.

3.2.4.1 Data repository

Data repositories are another strategy that could facilitate preservation of research data to ensure that researchers, scholars and other stakeholders can be assured of availability, accessibility, preservation and dissemination of content (Decman and Vintar 2013; and Yiotis 2008). An effective data repository is usually achieved through the collaborative works of librarians, information technologists, archivists, policymakers and research institute administration. On their part, research funders and publishers require agricultural research institutes to deposit research data in certified and credited repositories, in an effort to secure the reusability and long-term preservation of research data. RECODE Project Consortium (2014) agrees that obtaining accreditation or certification to appropriate standards is a way for ensuring both the quality of data repositories and of the quality assurance process.

Good data repositories are goldmines for agricultural research institutes because they bring in the benefits of open access as well as enhance sharing and reuse of research data. On the contrary, Parker (2012) points out that there are complexities around clarity of ownership, description and preservation formats of research data in data repositories which results in some discontentment among researchers. In this regard, Amorim et al. (2015); Bush (2009); and Fary and Owen (2013) state that if a clear and articulated RDM policy is in place, then issues of ownership, storage, formatting, description, networks, and software will clearly be spelled out to allow fluent RDM. Agricultural research institutes need to invest aggressively in data repositories in consultation and in collaboration with the government, agricultural research institutes and RDM stakeholders in order to change the landscape of scholarly communication across agricultural research institutes.

3.2.5 Data access

Open access refers to the practice of making peer-reviewed scholarly research and literature freely available online to anyone interested in reading it (European Commission 2016; and OECD 2007, 2004). It is imperative that open access movement (Berlin Declaration 2003; European Commission 2016; Higman and Pinfied 2015) have advocated for research data to be carefully preserved and made widely available through open access to enhance sharing, use and reuse. The Scholarly Publishing and Academic Resources Coalition (n.d) lists three key requirements for open access; first- availability, second- access and third- redistribution and reuse. In contrast, RECODE Project Consortium (2014) found out that most agricultural research institutes in the European Union focus on open access to publications rather than research data and therefore there is need to develop policies that allow the openness of research data but safeguarding the intellectual property rights, ownership while meeting the funders requirements. There are also institutional and community benefits provided by open data (Macdonald and Martinez-uribe 2010; Ball 2012; and NSF 2011) which include:

a) Visibility in terms of increased citation and usage, and greater public engagement;
b) Make new discoveries through faster impact, wider collaboration and increased interdisciplinary conversation; and
c) Comply with funder mandates.

In sum, open access is a means to advance knowledge, increase the benefits and return of investment in research and to foster innovation. Open access to research data from public funding should be easy, user-friendly and preferable internet based (Thanos 2010), but this will only be possible if policy, legal and technological dimensions are addressed, for instance,

technology must render physical and semantic barriers irrelevant, while policies and laws must address and supplant outdated legal jurisdictional boundaries.

3.2.6 Data use and reuse

The use of data collected in addressing emerging issues or being reused to find out whether a research establishes same conclusions or re-analyses of existing data to come up with powerful insights is imperative to RDM and agricultural research institutes (Lewis 2010; and Heidorn 2011). Improvement in technology, tools and communications have made research data easier to use and reuse. Since research data is data-rich, researchers have the opportunity to research into many aspects, including re-analysis of existing data, verification of results, minimization of duplication of efforts and acceleration of innovation, leading to improved production and attainment of good economic and social standing.

3.2.7 Sharing research data

There is need for agricultural research institutes to provide not only structures and policies for research data sharing, but services to support and educate researchers on concepts of data management and strategies for sharing data that can often be vital for the continuation of research (Karasti, Baker and Hakola 2006). Van den Eynden, Corti, Woollard, Bishop and Horton (2011) buttress the following as benefits of sharing research data:

a) Encourages scientific enquiry and debate;
b) Promotes innovation and potential new data uses;
c) Leads to new collaborations between data users and data creators;
d) Maximizes transparency and accountability;
e) Enables scrutiny of research findings;
f) Encourages the improvement and validation of research methods;
g) Reduces the cost of duplicating data collection;
h) Increases the impact and visibility of research;
i) Promotes the research that created the data and its outcomes;
j) Can provide a direct credit to the researcher as a research output in its own right and
k) Provides important resources for education and training.

In the same vein, Van den Eynden *et al.* (2011) describes various ways of sharing research data including:

a) Depositing them with a specialist data centre, data archive or data bank;
b) Submitting them to a journal to support a publication;
c) Depositing them in an institutional repository;
d) Making them available online via a project or institutional website and
e) Making them available informally between researchers on a peer-to peer basis.

The ease, with which digital research data can be stored, disseminated and made easily accessible online, means that many institutions should strive to share research data to enhance the impact and visibility of the research. The majority of the research funders in their research data sharing policy and mandates insist that publicly funded research data should, as far as possible, be openly available and encourage researchers to share data and outputs to the scientific community and stakeholders (American Psychological Association 2015).

In reviewing literature on data curation, Palathingal *et al*. (2015) concluded that a global trend on curating agricultural research data in the emergence of data-intensive research call for a well-designed technical infrastructure, trained human capital, policies and procedures at every stage of data curation and collaboration among agricultural research institutes in order to fill gaps at every stage of data curation.

3.3 Knowledge, skills and training requirements of RDM

Most extant studies have focused on the training of librarians, IT specialists, archivists and researchers to manage metadata, research data archives, data repositories, data curation, searching and retrieval, access, and web portals (Kuusniemi, Heino and Larmo n.d; Cox and Pinfield 2014; and Molloy and Snow 2012). The Society of College, National and University Libraries (SCONUL) (2015) advance the views of the Digital Curation Centre (DCC) and Research Libraries in United Kingdom on the knowledge and skills required of RDM staff in supporting RDM by highlighting the following:

1. RDM knowledge
   - ***a)*** The research data cycle and the stages of a research career in order to understand where support might be most needed;
   - ***b)*** RDM principles, including RDM planning, curation and preservation;
   - ***c)*** The discipline-specific nature of data; and
   - ***d)*** Open access and data sharing.

2. RDM skill
   - ***a)*** Advocacy, negotiation and diplomacy to work with researchers and other professional staff;
   - ***b)*** Guidance and training to support researchers in carrying out their responsibilities;
   - ***c)*** Advice and guidance to assist with necessary processes and procedures;
   - ***d)*** Understanding researcher requirements in order to anticipate and provide appropriate support;
   - ***e)*** Bibliometrics; and
   - ***f)*** Monitoring data reuse, citations and impact.

According to a study done by Henty (2014) on RDM competencies, the training needs of RDM occasioned by the influence of ICTs needs to address data generation, processing, preservation, dissemination, sharing and reuse. This view is consistent with the argument advanced by Taylor (2014) that skills acquired to curate, access, reuse, and share research data with the advent of new technologies remains one of the major challenges. Taylor continues to explain that researchers, librarians, archivist and IT specialists are deficient in knowledge and skill that pertains manipulation, interpretation and long term access to research data collected. Researchers often require training to enable them to acquire knowledge and skills needed to make their research data available and accessible or how to reuse data and incorporate data in their research process. Also, librarians, IT specialists and archivists require training on provision of research data services (RECODE Project Consortium 2014).

The right skills still need to be recruited and developed, and this would only be possible if the RDM stakeholders have a good understanding of the niche areas that need to be occupied. Therefore, agricultural research institutes should engage in professional development activities for RDM through workshops, conferences, seminars, as well as formal training programs and curricula that enable the gradual development of research data-scientists, RDM skills and knowledge. In the same perspective, RDM advocacy, awareness and data literacy should also be considered as imperative in advancing skills and competences in RDM. However, Lyon (2012) and Lewis (2010) propose that innovative approaches are needed to address the significant knowledge and skill gaps, data literacy and training in RDM especially in the areas of data curation, open access repositories and research data services.

3.4 Information and Communication Technology (ICT) infrastructure for RDM

The digital revolution is transforming the way scientific research is conducted. Henty (2014) illuminated that the growing contribution of ICTs to research has excited researchers the world over as they invest in new ways of conducting research and enjoy the benefits of more

sophisticated computers and communications systems that support measurement, analysis, modeling, simulation, collaboration and publishing. It is important in today's research environment for researchers to embrace the use of ICTs in order to effectively collect, analyse, preserve, share and manage research data while at the same time being able to access multi-scale, multi-discipline and multi-national research data. Australian government (n.d) asserts that development in ICTs is revolutionizing the research sector by setting up high-speed networks, web portals, metadata, and data repositories bringing in substantial potential benefits in data generation, analysis, manipulation, sharing and reuse. Amorim et al. (2015) emphasize the need for compatibility of data repositories, metadata, security systems, data management systems, search mechanisms and community acceptance as they are central to RDM.

A survey conducted by Maru (2004) on ICT in agricultural research and development in Sub-Saharan Africa found out that National Agricultural Research System (NARS) have major gaps and weaknesses in ICT infrastructure, such as

a) Capability, including skills and training;
b) Content, including generation and management;
c) Capital, with focus on funding;
d) Connectivity, not only physical but the ability to access information buy individual and user community;
e) Organisational or institutional culture; and
f) Conceptual framework (legal and policy framework) related to ICT.

Furthermore, Coherence in Information for Agricultural Research for Development (CIARD) (2012) did a case study on Kenya Agricultural Information Network (KAINet) and established that the use of Web 2.0, You Tube and Social Networking to enhance visibility and exchange of research outputs, including metadata, has not been widely implemented and embraced by most of Kenya's agricultural research institutes. In addition, Physical security, network security and security of computer system and files all need to be considered to ensure security of data and prevent unauthorized access, changes to data, disclosure or destruction of data (Princeton University 2017). Data security arrangements need to be proportionate to the nature of the data and the risks involved. Data security may be needed to protect intellectual property rights, commercial interests, or to keep personal or sensitive information safe (Van den Eynden et al. 2011). A qualitative study by Piennar (2010) conducted at the University of Pretoria on RDM practices revealed that ICT infrastructure such as web portals, institutional repositories, networks, hardware and software for RDM were not fully developed to facilitated full operationalization of RDM.

3.5 Collaborative partnerships influencing RDM

Collaboration within institution and among institutions is necessary for the sharing of research data, and for creating and sustaining public-private partnerships among research institutes, and partners (Pinfield, Cox and Smith 2014; Erway and Rinehart 2016; and Flores et al. 2015). Bracke (2011) illuminates that in the long run, however, RDM potential will only be tapped if the many actors in data creation, management and use are able to develop collaborations to build shared infrastructure and to develop and implement best RDM. Moreover, agricultural research institutes should uphold good governance that would promote collaboration within and across research areas, nationally and internationally, and ensure the effective establishment, operation and management of research data infrastructure. Collaborative research networks in agricultural research institutes create a significant amount of new data and it is imperative that this data is well managed to ensure that it is secure, discoverable, accessible, useable and re-usable (Gibson and Gross 2013).

Pinfield, Cox and Smith (2014) interviewed 26 respondents with regard to collaborative partnerships. The findings revealed that collaborative partnerships on one hand have benefits

such as metadata exchange, sharing, and reuse of research data and on the other hand, challenges such as lack of teamwork and policies governing collaborative partnerships. Australian Government (n.d) calls for a national collaborative approach to investment in research data infrastructure in order to reduce duplication, enhance economic use of resources, and optimize research outcomes and benefits. In addition, appropriate access arrangements and agreed standards will facilitate collaboration, fostering multi-disciplinary research uses for existing data, enabling researchers to address emerging problems in new ways.

4.0 RDM initiatives and challenges in agricultural research institutes

RDM in developed countries such as United Kingdom (UK), United States of America (USA), Australia and Canada have made great advancements (Lewis 2010; NSF 2007; and Henty 2014).

4.1 RDM perspective in United Kingdom

Large-scale management of the research data emerged over forty years ago in Europe when the UK Data Archives was established to manage paper-based surveys and other data outputs. This has been given impetus by the growth of digital research data and growing interest in long-term preservation, curation, and storage of research data for reuse and sharing (Lewis 2010). In the year 2001, the UK government funded e-Science Core Program, administered by the Engineering and Physical Sciences Research Council (EPSRC) on behalf of Research Councils UK to establish infrastructure, middleware and documentation to facilitate wide uptake of RDM (Lewis 2010; and Hey and Trefethen 2003). Furthermore, e-Science Core Program also supported demonstrator projects to enable researchers to understand the scope, capability, and implications of e-research projects and the need to manage data that was generated forthwith with a focus on areas (Lewis 2010) such as:

- Data-intensive: generating and often using large volumes of data;
- Collaborative: involving researchers across multiple institutions and transnational limitations;
- Grid-enabled: using high-capacity network and middleware.

According to Lord and Macdonald (2003), the Joint Information Systems Committee (JISC) commissioned a report on the curation of e-Science data and together with the e-Science Core Program, highlighted the role of the Digital curation centre with recommendations about the need to develop national capacity and capability to handle RDM. As can be seen, RDM featured prominently in e-Science Core Program and JISC in the UK which propelled the significance of Data Curation, (Data Curation Centre, DCC) lifecycle model (was a key recommendation in the JISC), technical infrastructure, legal issues, and human capabilities.

4.2 RDM perspective in United States of American (USA)

The growth of digital research has seen the emergence of data-intensive and collaborative research leading to the establishment of the National Science and Technology Council Committee (NSTCC) in the USA and the e-Infrastructure Reflection Group in the European Union to advise on capability, capacity and infrastructure in data management (Van den Eynden *et al*. 2011). These developments have increased investment in data management (Lewis 2010). For example, the USA- NSF has invested funds and cyber-infrastructure for research data curation through DataNet programme (NSF 2007). The introduction of DataNet program was as a result of the introduction of data management plan requirements by the NSF which could result in an effective and efficient RDM and for this reason drew attention to the need for data management infrastructure, both in terms of hardware, human and policy support (Halbert 2013; and NSF 2007). The NSF mandate was neither unprecedented nor an isolated intervention. The National Institutes of Health (NIH) had implemented the first mandate in 2003, requiring researchers to comply with data sharing and data management practices (NIH 2003). Other federal agencies, for example, the National Endowment for the Humanities, adopted a

requirement for data management plan that explicitly emulated the NSF requirement (National Endowment for the Humanities 2013).

To this end, it can be noted that the USA government has laid down strategies to create a comprehensive framework of transparent, evolvable and extensible policies, infrastructure, management and organisational structures that provide reliable and effective access, reuse and sharing of research data.

4.3 RDM perspective in Australia

Australia has also moved with relative alacrity to develop data management of e-research and has set up the Australian National Data Services (ANDS) funded by the Australian government through the National Collaborative Research Infrastructure Strategy (NCRIS) (ANDS Technical Working Group 2007). In this regard, ANDS core purpose is to make Australia's research data assets more valuable for researchers, agricultural research institutes and the nation. Henty (2014), in support of this view opines that Australian agricultural research institutes and researchers are keen to have greater access, sharing, and reuse of research data resulting in greater efficiency in RDM. In light of newly developed strategy, "The 2011 Strategic Roadmap for Australia Research Infrastructure", the Australian government has made significant investments in research data infrastructure to facilitate collection, generation, manipulation, curation, access and dissemination (Australian Government n.d). Carrick (2014) points out that RDM in Australia is an essential component of all research leading to the establishment of, and sharing of 'Australian Code for the Responsible Conduct of Research', jointly developed and issued by the National Health and Medical Research Council (NHMRC), the Australian Research Council (ARC), and Universities Australia. The code assigns researchers and their institutions the responsibility of addressing ownership, storage and retention, access to, and sharing of research data.

4.4 RDM perspective in Canada

In Canada, RDM services have become a high priority for government agencies and post-secondary institutions in recent years. While it has lacked coherent national strategies for developing the digital infrastructure required for e-research, such as those in Australia or the UK, still there remains a growing expectation for sound RDM (Whitehead and Bourne-Tyson n.d). Due to different aspects of digital infrastructure being operated separately at different government levels and not as part of a cohesive whole at the national and institutional levels, the Canadian government funded three main federal research granting council known as the 'Tri-Agencies' with the sole purpose of strengthening RDM in Canada and maintain Canada's research excellence (Government of Canada 2016). The Tri-Agencies (Whitehead and Bourne-Tyson n.d; and Government of Canada 2016) are:

a) The Canadian Institutes of Health Research;

b) The Natural Sciences and Engineering Research Council of Canada; and

c) Social Sciences and Humanities Research Council (SSHRC).

According to Fry, Doiron, Létourneau, Perrier, Perry et al. (2017), the Tri-Agency Statement of Principles on Digital Data Management heightens the need for a collaborative national perspective on RDM that has been missing in Canada. Under its auspices, the statement outlines the agencies' overarching expectations for RDM and the role of researchers, agricultural research institutes, research communities, and research funders in supporting data management (SSHRC 2015). In Addition, the Canadian government through the Tri-Agencies promotes and supports research, research training, knowledge transfer and innovation within Canada. Like other developed countries, the Canadian government and SSHRC (Government of Canada 2016; and SSHRC 2015;) maintains that in promoting access to research results, the Tri-Agencies aspire to advance knowledge, avoid research duplication and encourage reuse,

maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. In this respect, the Canadian government and research funders are becoming increasingly aware of the value of research data, the importance of fostering reuse of research data and the need for policies to enable excellence in RDM (Government of Canada 2016; and Sewerin 2015).

4.5 RDM perspective in South Africa

South Africa is leading the cluster of African countries in embracing RDM (Van Deventer and Piennar 2015). Some research councils and institutes, and academic and research libraries in South Africa have initiated programs towards the realization of RDM. For example, the University of Cape Town (UCT) has established e-Research centre to work and partner with researchers in finding IT solution for their research work while the University of South Africa (UNISA) has completed investigation into RDM as part of the plan to establish data management (Macanda, Rammutloa and Bezuidenhout 2015). The University of Pretoria, Stellenbosch and Witwatersrand are at different planning and implementation stages (Van Wyk and Van der Walt 2014; Van Deventer and Piennar 2015). An investigation on “Research Data Management in South Africa” by Kahn, Higgs, Davidson and Jones (2014) found that in South Africa, a number of data repositories have been established to manage research data. They include South African National Park, National Health Information Repository and Data Warehouse, and Data Intensive Research Initiative of South Africa (DIRISA). Lötter (2014) and Fernihough (2011) affirm that DIRISA is one of the initiatives aimed at promoting RDM in the country.

Van Deventer and Piennar (2015) postulates that the Council for Scientific and Industrial Research (CSIR), DIRISA and National Integrated Cyberinfrastructure System (NICIS) are setting systems in place in terms of data curation, policies, technical infrastructure, and human capability to enable management, reuse, and sharing of research data. In the same vein, the National Research Foundation (NRF) - South Africa, as the leading government research funding agency, has been involved in many initiatives to allow the sharing of research outputs, datasets research support and knowledge networking databases which contribute to knowledge generation for the support and promotion of research development (NRF 2015). Studies done by Kahn *et al*. (2014); Lötter (2014); and Van Deventer and Piennar (2015) on RDM landscape in South Africa, shows that there are high levels of awareness of RDM in research and academic institutions with notable activities of Network of Data and Information Curation Communities (NeDICC) in most agricultural research institutes. NeDICC aims to promote the development and use of research data including curation standards and practices of agricultural research institutes to ensure the long term preservation and accessibility of digital research outputs (UCT 2016). Kahn et al. (2014) nevertheless singled out issues of leadership, policy and skills as areas that require the most urgent attention.

4.6 RDM perspective in Kenya

In Kenya, there is some attempt, albeit limited, to promote RDM especially in the health and migration sector (Jao , Kombe, Mwalukore, Bull, Parker et al. 2015; Family Health International-Kenya 2005; Olum 2013). Olum (2013) posits that despite Kenya having adequate migration data in various institutions, the data is not sufficiently coordinated, shared, analysed or disseminated. Olum further observed that many of the government and civil institutions have limited capacity, resources and facilities for collection, analysis, use and reuse, and sharing of migration data hence making access and use of migration data difficult. Furthermore, the health sector in Kenya is making strides in embracing the sharing of research data albeit at a rate of limited rapidity. Lairumbi, Parker, Fitzpatrick and Mike (2011) opined that health research data sharing in Kenya is vital among health researchers for example, sharing research data on malaria or HIV/AIDS could enhance clinical vaccine trials and advance breakthroughs in the

health sector. However, sharing health research data is not fully embraced, leading to underutilization of research data.

World Agroforestry Centre based in Kenya organized RDM training for agroforestry research scientists whose aim was to encourage necessary allocation of resources such as, skilled personnel and technical infrastructure for data management (World Agroforestry Centre 2002). Equipped with the RDM training manual, the World Agroforestry Centre made an incredible step in introducing RDM into the agriculture sector. The assessment by scholars (Chisenga 2012; Mugata 2014; Alila and Atieno 2006; World Agroforestry Centre 2002) on RDM in Kenya established that various institutions, specifically agricultural research institutes have rich valuable research data that needs to be managed effectively in order to enhance the institutes mandate, increase agricultural productivity and enhance food security. Mugata (2014) asserts that accessing agricultural research data in Kenya is not easy for researchers and other stakeholders in agriculture due to limited institutional skills, legal framework, infrastructure and strategies that support RDM. However, to facilitate agricultural related content accessibility, visibility and sharing in Kenya, Kenya Agricultural Information Network (KAINet) was established (Mugata 2014). KAINet aims at building a common and freely accessible information system for the generation, collection, processing, preservation and dissemination of agricultural research data and information.

5 Discussions

In analyzing and evaluating literature review on RDM the following issues emanates.

Erway (2013); and Higman and Pinfield (2015) point out that failure to establish legal and policy protocols for RDM is likely to diminish the potential for valuable research outputs to be made available for access, sharing and reuse .For example, Agricultural research institutes are increasingly getting involved in data-intensive research projects that cut across disciplinary borders and also involve communities of researchers participating in large-scale collaborations. In this respect, there is a need to develop legal frameworks, policies and regulations on RDM to facilitate systems and services in enabling research data to be managed, accessed, shared, reused and secured. Fitzgerald and Pappalardo (2007) assert that RDM occurs in legal and policy contexts and the principal areas of law that cover RDM includes copyright, moral rights, patents, confidentially, contract, and privacy. In addition, Erway (2013); and Jones, Pryor and Whyte (2013) emphasize that policies and regulations at every stage of research lifecycle ensure that consistent data management standards and quality are maintained in order to foster management of and access to research institute's intellectual assets and also provide uniform requirements to facilitate data understandability and sharing among research data stakeholders.

According to Higgins (2008), data curation consists of a range of activities and processes focused on capture, appraisal, description, preservation, access, reuse and sharing which adds value to research data throughout its lifecycle. United Nations Environmental Program (UNEP) (2015) claims that in developing countries, agricultural research data existing in different formats or software versions scattered across servers, computers, storage devices or other filing systems, which makes it difficult to change into modern formats or versions and in the process research data is lost. In this regard, there is need for standardized format of data curation to allow uniformity. Concerning funding, Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data (2016) asserts that the sustainability of RDM represents a challenge within the existing funding structures especially research institutes in developing countries and at the core of this particular challenge are issues related to the eligibility of funding for research to allow generation of research data for curation. Moreover, there is a lack of awareness about the importance of metadata among the scientific community and therefore there is a need for research institutions to organize RDM literacy programs to assist researchers and other RDM stakeholders to understand, prepare and use metadata

necessary to enable the discovery, preservation, and reuse of their data (Tenopir et al. 2011). Concerning preservation, RECODE Project Consortium (2014) emphasizes that policies and regulations regarding retention periods and where to deposit research data when researchers leave the research institute, should clearly be stipulated.

On the same breath, European Commission (2016) posits that fuller and wider access to research data is important as it helps to build on previous research results, encourage collaborations and help avoid duplication of effort, speed up innovation, and involve citizens and society in the scientific process. Access to Kenya's agricultural research data faces various challenges as enumerated by Muinde and Gorman (2009) that include: social-cultural (non-visionary leadership), lack of ICT infrastructure, legal, policy and institutional frameworks, and capacity building programs. Muinde and Gorman (2009); and Chisenga (2012) point out that there is a necessity for government to fund, plan and prioritize resource allocation for research to ensure that there is local content which is visible, accessible and sharable online to facilitate information flow.

On knowledge, skill and training required for RDM, literature indicated that a study done by Schmidt and Shearer (2016) on librarians' competencies profile for research data management, guided by interpretive paradigm, enumerated the knowledge that RDM staff should posses must include: knowledge of repositories, data manipulation, data discovery mechanisms, funders' policies and requirements, data centers, data publication requirements of journals, sharing and access, data citation and referencing, metadata standard and schemas among others. In addition, Fary and Owen (2013); and Creamer, Morales and Crespo (2012), in their studies on RDM skills and competencies listed the skills relevant to RDM which included: storage, data migration, networking, legal, financial, security, metadata creation and assignment, scholarly data communications, and preservation. Kennan (2016) interviewed 25 data professionals in Australian scientific research organisation and found out that the most common set of skills required were: interpersonal skills, data specific knowledge and skills, and metadata. In this regard, it would seem knowledge and skills possessed by researchers and other RDM stakeholders in Kenya's agricultural research institutes were inadequate. However, Lötter 2014); and Patrick et al. (2013) pointed out that RDM training should be undertake through capacity building workshops; in-house training and mentorship of research data curators as well attending conferences and seminars.

As far as ICT infrastructure for RDM is concern, literature revealed that while cyberinfrastructure has been revolutionizing digital research, a comprehensive framework for capturing, organizing, preserving, and making research data available and usable has not been created, further access to adequate ICT tool and equipment has not be in tandem with the training of researchers and other data professionals on how to utilize them effectively in RDM (Witt 2008). In relation to ICT skill, Amorim et al. (2015) emphasize that there is a need for compatibility of research data security with data repositories, metadata, security systems, data management systems, and search mechanisms in order to enhance privacy of research data. Moreover, data security is important for protecting intellectual property rights, commercial interests, or to keeping personal or sensitive information safe (Van den Eynden et al. 2011). Therefore, Maru (2004) in a study sums up gaps and weaknesses of ICT used in Kenya's agricultural research institutes to be as follows: Capacity (including infrastructure and skill); Content (including generation and management); Capital (with the focus on funding not only ICT also capacity building); connectivity (not only physical but the ability to access information by individual and user community); and collaboration (within and across research institutes at national, regional and international).

Collaborative partnerships are important catalysts in research project as literature indicated that collaboration between researchers and the public contributes significantly to increased productivity, quality of agricultural products, and diversified crops and livestock (KALRO- Tea

Research Institute 2016). Meanwhile, Pinfield, Cox and Smith (2014) conducted interviews with 26 respondents regarding collaborative partnerships. The findings revealed that collaborative partnerships have benefits such as metadata exchange, exchange RDM human capacity, sharing and reuse of research data on one hand and on the other hand challenges such as lack of teamwork and policies governing collaborative partnerships undermine it. Jahnke and Asher (2012) notes that there is a great need for more effective collaboration tools, as well as online spaces that support the volume of data generated and provide appropriate privacy and access controls.

Therefore, literature revealed that despite absence of legal framework in the agricultural research institutes, individual institutes had RDM policies guiding researchers on research data capture, appraisal, description, preservation, access, reuse and sharing. Further indication is the lack of coordinated RDM strategies in the research institutes led to loss of data and difficulty in accessing, reusing and sharing research data. Furthermore, literature revealed that there was scarcity of RDM knowledge and skill in the agricultural research institutes hence for example, limiting the utilization of research data in Kenya's agricultural research institutes. Generally literature indicated that ICT infrastructure for capturing, appraisal, description, preservation, access, sharing and security of data was inadequate to facilitate RDM. Collaborative partnership enhanced management, access and sharing of research data. Generally, the finding from literature indicated that support from government, research institutes, and collaborative partners could enhance management, access, reuse and sharing of agricultural output (Pinfield, Cox and Smith 2014; Cox, Verbaan & Sen 2012; Qin 2013; Mossink, Bijsterbosch, & Nortier 2013; Grebmer and Spielman 2004; and Lewis 2010).

In a nutshell, reviewed literature in this paper gives content analysis on background information on RDM issues and plays a crucial role in preparing and orientating the researcher with regard to the raging debates taking place in RDM field. More or less it is descriptive in form as Cooper (2011:20) alludes that reviewed literature integrate what others have done and said, criticize previous scholarly works, build bridges between related topics, and identify the central issues in a field. On the other hand reviewed literature cannot provide current trends and investigations to RDM in the wider institutional context, the ongoing drivers for RDM activities and factors influencing the shape of RDM development. However, there is need to undertake empirical studies in RDM in order to provide current detailed synthesis, sufficient criticisms, clear methodology, up-to-date data /information, and further reveal areas of concern that need investigation.

Conclusion

The literature reviewed revealed that RDM have been recognized worldwide, albeit to different extents with more pressure globally on agricultural research institutes embrace RDM to facilitate research data capture, appraisal, preservation, access, sharing and reuse. Furthermore, literature reviewed revealed that for the successful establishment of RDM in agricultural research institutes, the establishment of a formal data governance structure to address RDM issues, enactment of RDM legal, policies and regulations; capacity building programs and plans, incentivisation of researchers; a sound technical infrastructure and collaborative partnerships in boosting relationships among research institutes.

Besides, literature reviewed revealed that RDM in Kenya , for example, is given little attention as attested by the limited documentation or publications of the research that is generated by these institutions, leading to poor management, limited access of such research data, duplication of the research, poor sharing and reuse of the research data (Jao et al. 2015; Family Health International-Kenya 2005; The World Agroforestry Centre 2012; Alila and Atieno 2006; Mugata 2014). This review provide frameworks particular of specific to developing countries by providing evidence of specific challenges that such countries contend with (Ng'eno

2018). Moreover, few literature review on RDM have been done in Africa. The current literature review is therefore significant in contributing to the scholarly research on RDM in developing countries such as Kenya (Ng'eno 2018).

Acknowledgement

I am indebted to my supervisor Prof. Stephen Mutula for his devotion and admirable guidance when writing this article. He gave selflessly of himself in providing technical advice and supervision in making sure the article contributed to knowledge. My gratitude and appreciation to University of KwaZulu-Natal and employer, Moi University, for granting me the opportunity and study leave respectively to undertake this PhD program. In addition, the literature review is revised from my PhD Titled 'Research Data Management in Kenya's Agricultural Research Institutes'- University of KwaZulu-Natal, 2018(the data collection intruments for the PhD can be replicated in related studies).

References

Alila, P. and Atieno, R. 2006. Institute of Development Studies Agricultural Policy in Kenya: Issues and processes. *A paper for the future agricultures consortium workshop, Institute of Development Studies, 20-22 March 2006*. Viewed 25 April 2016, from: *http://www.fao.org/fileadmin/user_upload/fsn/docs/Ag_policy_Kenya.pdf*

American Psychological Association. 2015. Data sharing: principles and considerations for policy development. Viewed 5 October 2016, from: *https://www.apa.org/science/leadership/bsa/data-sharing-report.pdf*

Amorim, R.C., Castro, J.A., Silva, J. R. and Ribeiro, C. 2015. A comparative study of platforms for research data management: interoperability, metadata capabilities and integration potential. Viewed 13 May 2016, from: *http://link.springer.com/chapter/10.1007/978-3-319-16486-1_10#page-1*

Australian Government. n.d. The Australian research data infrastructure strategy. Viewed 5 August 2016, from: *https://docs.education.gov.au/system/files/doc/other/the_australian_research_data_infrastructure_strategy.pdf*

Australian National Data Service. n.d. Data re-use. Viewed 27 October 2016, from: *http://www.ands.org.au/working-with-data/enabling-data-re-use/data-re-use*

Austrian National Data Services (ANDS) Technical Working Group. 2007. Towards the Australian data commons: a proposal for an Australia national data service. Department of education, science and training. Viewed 26 May 2016, from: *http://www.ands.org.au/__data/assets/pdf_file/0009/386982/ardc_eif_annual_report_2011-12.pdf*

Ball, A. 2012. Review of data management lifecycle models. Viewed 20 July 2016, from: *http://opus.bath.ac.uk/28587*

Berlin Declaration. 2003. Berlin declaration on open access to knowledge in the sciences and humanities. Viewed 25 April 2016, from: *http://www.fu-berlin.de/sites/open_access/Veranstaltungen/oa_berlin/poster/Berlin-Dclaration_Simone-Rieger_MPIWG.pdf*

Bracke, M.S. 2011. Emerging data curation roles for librarians: a case study of agricultural data. *Journal of Agricultural and Food Information,* 12 (1), 65-74.

Bush, V. 2009. Research data management. Viewed 14 May 2016, from: *https://net.educause.edu/ir/library/pdf/ers0908/rs/ers09087.pdf.*

Carrick, A. 2014. Research data management: guidelines and planning for researchers. Viewed 22 May 2017, from: *https://www.swinburne.edu.au/media/swinburneeduau/research/docs/pdfs/Research_Data_Management_Guidelines.pdf*

Chisenga, J. 2012. *Opening access to agricultural sciences and technical information in Ghana, Kenya and Zambia.* Presented at SCECSAL XXth conference hosted by Kenya Library Association (KLA) on 4th-8th July, 2012. Venue LAICO REGENCY HOTEL, Nairobi, Kenya. Viewed 4 March 2017, from: *https://core.ac.uk/download/pdf/11890652.pdf*

Coherence in Information for Agricultural Research for Development (CIARD). 2012. Case studies: Kenya Agricultural Information Network (KAINET). Viewed 30 March 2016, from: *http://www.ciard.info/sites/default/files/RAF%20(Kenya).pdf*

Cooper, H. M. 2011. *Research synthesis and meta-analysis: a step by step approach*. 4th (ed). London: Sage.

Cornell University Library. n.d. Metadata and describing data. Viewed 6 June 2016, from: *http://data.research.cornell.edu/content/writing-metadata*

Cox, A.M. and Pinfield, S. 2014. Research data management and libraries: current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316.

Cox, A., Verbaan, E. and Sen, B. 2012. Upskilling liason librarians for research data management. Viewed 27 October 2016, from: *http://www.ariadne.ac.uk/issue70/cox-et-al*

Creamer, A., Morales, M.E. and Crespo, J. 2012. An assessment of needed competencies to promote the data curation and management librarianship of health sciences and science and technology librarians in New England. *Journal of eScience Librarianship*, 1(1), 18-26.

Data Curation Centre. 2004. Planning preservation. Viewed 4 July 2016, from: *http://www.dcc.ac.uk/digital-curation/planning-preservation*

Decman, M. and Vintar, M. 2013. A possible solution for digital preservation of e-government: a centralised repository within a cloud computing framework. *Aslip Proceedings: New Information Perspectives,* 65, 406-424.

Dora, M. and Kumar, H. A. 2015. Managing research data in academic institutions: role of libraries, 2015. In CALIBER, 2015, Shimla, March 12-+14, 2015 [Conference paper]. Viewed 18 September 2017, from: *http://eprints.rclis.org/24911/1/50.pdf*

Eaker, C. 2016. Selection and appraisal of digital research datasets. In L. Kellam and K. Thompson (eds.). *Databrarianship: The academic data librarian in theory and practice. American Library Association.* Viewed 4 October 2016, from*: http://trace.tennessee.edu/cgi/viewcontent.cgi?article=1018andcontext=utk_libpub*

Erway, R. 2013. Starting the conversation: university-wide research data management policy. Dublin, Ohio: OCLC Research. Viewed 1 May 2016, from*: http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf*

Erway, R. and Rinehart, A. 2016. *If you build it, will they fund? making research data management sustainable*. Dublin, Ohio: OCLC Research.

European Commission. 2016. Guidelines on open access to scientific publications and research data in horizon 2020: version 2.1. Viewed 20 April 2016, from: *http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf*

Family Health International-Kenya. 2005. Strengthening research and data management: new efforts at the Kenya division of reproductive health. Viewed 29 March 2016, from: *http://pdf.usaid.gov/pdf_docs/Pnadk371.pdf*

Fary, M and Owen, K. 2013. Developing an institutional research data management plan service. Viewed 10 March 2016, from: http *https://net.educause.edu/ir/library/pdf/acti1301.pdf*

Fellous-Sigrist, M. 2015. What are research data? Viewed 1 March 2017, from*: https://www.ucl.ac.uk/tobedeleted-sites/dis-old-June16/research/RDM_presentation.pdf*

Fernihough, S. 2011. E-research: an implementation framework for South African organisation (Unpublished master's research report, University of South Africa). Viewed 26 May 2016, from: *http://hdl.handle.net/10500/4474*

Fitzgrald, A. and Pappalardo, K. 2007. Building the infrastructure for data access and reuse in collaborative research. Viewed 4 June 2017, from: *https://eprints.qut.edu.au/8865/1/8865.pdf*

Fitzgerald, A., Pappalardo, K. and Austin, A. 2008. Practical data management: a legal and policy guide. Viewed 5 June 2017, from: *https://eprints.qut.edu.au/14923/1/Microsoft_Word_-_Practical_Data_Management_-_A_Legal_and_Policy_Guide_doc.pdf*

Flores, J.R., Brodeur, J.J., Daniels, M.G., Nicholls, N. and Turnator, E. 2015. Libraries and the research data management landscape. Viewed 3 December 2015, from: *http://www.clir.org/pubs/reports/pub167/RDM.pdf*

Fry, J., Doiron, J., Létourneau, D, Perrier, L. and Perry, C. et al. 2017. Research data management training landscape in Canada: a white paper. Viewed 21 March 2017, from: *https://portagenetwork.ca/wp-content/uploads/2017/02/Portage_ Training_White_Paper_EN.pdf*

Gibson, D. and Gross, J. 2013. Research data management in a collaborative network. Research global. Issue 33, pp 12-15. Viewed 30 June 2016, from: *http://ro.ecu.edu.au/ecuworks2011/586*

Government of Canada. 2016. Tri-agency statement of principles on digital data management. Viewed 23 March 2017, from: *http://www.science.gc.ca*

Grebmer, K.V. and Spielman, D.J. 2004. Public-private partnerships in agriculture research: an analysis of challenges facing industry and the consultative group on international agricultural research. Washington, D.C: International Food Policy Research Institute.

Halbert, M. 2013. Prospects of research data management. In *DataRe.* r*esearch data management: principles, practices, and prospects.* Washington, DC: Council on Library and Information Resources. 1-15. Viewed 20 May 2016, from: *http://creativecommons.org/licenses/by-sa/3.0/*

Heidorn, P.B. 2011. The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51: (7-8), 662-672.

Henty, M. 2014. Dreaming of data: the library's role in supporting e-research and data management. Viewed 3 January 2016, from: *https://digitalcollections.anu.edu.au/bitstream/1885/47617/5/henty_alia_08.pdf*

Hey, T. and Trefethen, A. 2003. The data deluge: an e-science perspective. In F. Berman, G.C. Fox, and T. Hey, (eds), *Grid computing: making the global infrastructure a reality.* New York: Wiley.

Higgins, S. 2008. The DCC curation lifecycle model. *The International Journal of Digital Curation*, 3(1). Viewed 24 March 2016, from: *http://www.ijdc.net/index.php/ijdc/article/ viewFile/69/48*

Higgins, S. 2012. The lifecycle of data management. In G. Pryor. (ed.). *Managing research data*. London: Facet Publishing.

Higman, R. and Pinfield, S. 2015. Research data management and openness: the role of data sharing in developing institutional policies and practices. *Program: Electronic Library and Information System,* 49 (4), 364-381.

Ingram, C. 2016. How and why you should manage your research data: a guide for researcher. Viewed 2 February 2016, from: *https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data*

International Standards Organisation (ISO). 2001. *ISO 15489-1:2001 information and documentation: records management standard. part 1: General*. Geneva: International Organisation for Standardization.

Jahnke, L. M. and Asher, A. 2012. The problem of data: data management and curation practices among university researchers. In. L. Jahnke, A. Asher and S. Keralis. *The problem of data*. Washington, DC: Council on Library and Information Resources. Viewed 3 September 2016, from: *https://www.clir.org/pubs/reports/pub154/pub154.pdf*

Jao, I.., Kombe, F., Mwalukore, S., Bull, S., Parker, M., Kamuya, D., Molyneux, S. and Marsh, V. 2015. Research stakeholders' view on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations. Viewed 20 April 2016, from: *http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135545*

Jones, S., Pryor, G and Whyte, A. 2013. *How to develop research data management services: a guide for HEI's.* Edinburgh: digital curation centre. Viewed 23 September 2015, from: *http:// www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services*

Kahn, M., Higgs, R. Davidson, J. and Jones, S. 2014. Research data management in South Africa: how we shape up. *Australian Academic and research Libraries*, 45(4), 296-308. Viewed 30 March 2016, from: *http://dx.doi.org/10.1080/00048623.2014.951910*

Karasti, H., Baker, K. and Hakola, E. 2006. Enriching the notion of data curation in e-science: data managing and information infrastructure in the Long Term Ecological Research (LTER) network. Viewed 3 May 2016, from: *http://www.worldpece.org/sites/default/files/57-227-1-PB.pdf*

Karick, A. 2014. Research data management: guidelines and planning for researchers. Viewed 4 October 2017, from: *https://www.swinburne.edu.au/media/swinburneeduau/ research/ docs/ pdfs/Research_Data_Management_Guidelines.pdf*

Kennan, M. A. 2016. Data management: knowledge and skills required in research, scientific and technical organisations. Viewed 27 June 2017, from*: http://library.ifla.org/1466/1/221-kennan-en.pdf*

Kenya Agricultural and Livestock Research Organisation (KALRO)-Tea Research Institute. 2016. *Information for visitors*. Kericho, Kenya: Tea Research Institute.

Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data. 2016. Funding research data management and related infrastructures: knowledge exchange and science Europe briefing paper. Viewed 12 August 2016, from: *http://www.scienceeurope.org/wp-content/uploads/2016/05/SE-KE_Briefing_Paper_Funding_RDM.pdf*

Kruse, F. and Thestrup, J.B. 2014. Research libraries' new role in research data management, current trends and visions in Denmark. *Liber Quarterly,* 23(4), 310-335.

Kuusniemi, M. E., Heino, T. and Larmo, K. n.d. How to get started with research data management training services for the academic library? Viewed 16 April 2016, from: *http:// www.iss.it/binary/eahi/cont/76_Mari_Elisa_Kuusniemi_Full_text.pdf*

Lairumbi, G.M., Parker, M., Fitzpatrick, R. and Mike, E.M. 2011. Stakeholders understanding of the concepts of benefit sharing in health research in Kenya: A qualitative study. Viewed 24 March 2017, from: *https://bmcmedethics.biomedcentral.com/articles/10.1186/1472-6939-12-20*

Lewis, M.J. 2010. Libraries and the management of research data. In: E*nvisioning future academic library services*. Facet Publishing: London. 145-168.

Lord, P. and Macdonald, A. 2003. E-science curation report: 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. Twickenham, England: The Digital Archiving Consultancy Limited. Viewed 2 September 2015, from: *http://www.jisc.ac.uk/uploaded documents/e-scienceReportFinal.pdf*

Lötter, L. 2014. *Reflections on the RDM position in South Africa: development research data management services*. Cape Town, South Africa: Research Methodology Centre.

Lyon, L. 2012. The informatics transform: re: engineering libraries for the data decade. *International Journal of Digital Curation*, 7(1), 126-138.

Lyon, L., Ball, A., Duke, M. and Day, M. 2012. Community Capability Model Framework. Viewed 12 march 2016, from: *http://communitymodel.sharepoint.com*

Lyon, L., Patel, M. and Takeda, K. 2014. Assessing requirements for research data management support in academic libraries: introducing a new multi-faceted capability tool. Viewed 28 November 2015, from: *http://ozk.unizd.hr/proceedings/ index.php/lida/ article/view/121*

Macanda, M., Rammulloa, M., and Bezuidenhout, R. 2015. Research data management at University of South Africa (UNISA). Viewed 24 March 2017, from: *http://hdl.handle.net/ 10500/139007*

Macdonald, S. and Martinez-Uribe, L. 2010. A new role for academic librarians: data curation. *El Profesional de la Informacio'n*, 17(3): 273-280.

Maru, A. 2004. ICT/ICM in agricultural research and development: status in Sub-Saharan Africa. Viewed 3 June 2016, from: *http://www.fao.org/docs/eims/upload/ 210126/ ICT_in_ARD_Regional_Status_2004_FARA.pdf*

Molloy, L. and Snow, K. 2012. The data management skills support initiatives: synthesizing postgraduate training in research data management. *International Journal of Digital Curation*, 7(2), 101-109.

Mossink, W., Bijsterbosch, M. and Nortier, J. 2013. European landscape study of research data management: support infrastructure models for data management. Viewed 15 March 2016, from: *http://www.clarin.nl/sites/default/files/SIM4RDM% 20landscape% 20report_1.pdf*

Mugata, R. 2014. KAINeT initiative opening up access to agricultural research information in Kenya. Viewed 22 April 2016, from: *http://www.ciard.info/news-and-events/blog/kainet-initiative-opening-access-agricultural-research-information-kenya*

Muinde, F.N.N. and Gorman, G.E. 2009. Barriers to open access to scientific information in Kenya with particular reference to agricultural information. Viewed 26 August 2016, from: *http://www.ifla.org/annual-conference/ifla75/index.htm*

National Archives and Records Administration (NARA). 2007. Strategic directions: appraisal policy. Viewed 23 June 2016, from: *http://www.archives.gov/records-mgmt/initiatives/ appraisal.html*

National Endowment for the Humanities. 2013. Data management plans for NEH Office of digital humanities proposals and awards. Viewed 15 March 2017, from: *https://www.clir.org/pubs/reports/pub160/pub160.pdf*

National Institutes of Health (NIH). 2003. Final NIH statement on sharing research data. Viewed 30 March 2017, from: *http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html*

National Research Foundation (NRF). 2015. Databases. Viewed 15 February 2017, from: *http://www.nrf.ac.za/information-resources/databases*

National Science Foundation (NSF). 2007. Cyberinfrastructure vision for 21st century discovery. Viewed 6 July 2016, from: *http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf*

National Science Foundation (NSF). 2011. Digital research data sharing and management, No.NSB-11-79. Viewed 20 July 2016, from: *https://www.nsf.gov/nsb/publications/ 2011/nsb1124.pdf*

Niu, J. 2014. Appraisal and selection for digital curation. *International Journal of Digital Curation,* 9(2), 65-82.

Ng'eno, E.J. 2018*. Research data management in Kenya's agricultural research institutes.* Unpublished D. Phil. Thesis. KwaZulu-Natal, SA: University of KwaZulu-Natal, Pietermaritzburg, School of Social Sciences/Department of Information Studies.

Olum, G. H. 2013. *Enhancing migration data management in Kenya: assessment and recommendations.* Nairobi: ACP Observatory on Migration.

Organisation for Economic Co-operation and Development (OECD). 2004. OECD principles of corporate governance. Viewed 7 May 2017, from: *https://www.oecd.org/corporate/ca/corporategovernanceprinciples/31557724.pdf*

Organisation for Economic Co-operation and Development (OECD). 2007. OECD principles and guidelines for access to research data from public funding. Viewed 4 July 2016, from: *http://www.oecd.org/sti/sci-tech/38500813.pdf*

Palathingal, L., Dascalu, S., Harris, F.C. and Varol, Y. 2015. A brief of data curation literature. Viewed 20 May 2016, from: *https://www.cse.unr.edu/~fredh/papers/conf/143-absodcl/paper.pdf*

Palmer, C.L., Cragin, M.H., Heidorn, P.B. and Smith, L.C. 2007. Data curation for the long tail of science: the case of environment studies. Viewed 28 March 2016, from: *http://www.researchgate.net/publication/228656107*

Parker, R. 2012. What the library did next: strengthening our visibility in research support. Viewed 18 July 2016, from: *http://www.vala.org.au/direct-download/vala2012-proceedings/435-vala2012-session-1-parker-paper/file*

Patrick, M., Byatt, D., De Luca, F., White, W. Rumsey, S. and Wilson, A. J. 2013. Research data management training for support staff. A DaMaRO project survey. Viewed 5 July 2016, from: *http://damaro.oucs.ox.ac.uk/docs/Research_Data_Management Training _for%20_Support_Staff_-_a_DaMaRO_Project_Survey.pdf*

Pienaar, H. 2010. Survey of research data management practices at the University of Pretoria, South Africa: October 2009-March 2010. Viewed 28 March 2016, from: *http://www.ais.up.ac.za/profile/heila_pienaar/index.htm*

Pinfield, S., Cox, A.M. and Smith, J. 2014. Research data management and libraries: relationship, activities, drivers and influences, *PLoS ONE, 9(12)*. Viewed 25 March 2016, from: *http://eprints.whiterose.ac.uk/82499/1/PLOS%20ONE% 20final%20published% 20version.pdf*

Princeton University. 2017. Research data security. Viewed 11 October 2017, from: *https:// www.princeton.edu/ria/human-research-protection/data/*

Qin, J. 2013. Infrastructure, standards, and policies for research data management. Viewed 16 April 2016, from: *http://surface.syr.edu/cgi/viewcontent.cgi?article=1164andcontext=istpub*

Ray, J. M. 2014. Introduction to research data management. In. J. M. Ray. (ed). *Research data management: practical strategies for information professionals.* West Lafayette, Indiana: Purdue University Press.

RECODE Project Consortium. 2014. Policy recommendations for open access to research data. Viewed 3 August 2016, from: *http://www.recodeproject.eu*

Schellenberg, T.R. 1956. *The appraisal of modern public records (Bulletins of the National Archives No. 8)*. Washington, DC: Government Printing Office.

Schmidt, B. and Shearer, K. 2016. Librarians' competencies profile for research data management. Viewed 26 May 2017, from: *https://www.coar-repositories.org/files/ Competencies-for-RDM_June-2016.pdf*

Scholarly Publishing and Academic Resource Coalition. n.d. Institutional repositories and open access-the future of scholarly communication. Viewed 4 October 2016, from: *http:// www.budapestopenaccessinitiative.org/pdf/David_Prosser.pdf*

Sewerin, C. 2015. Research data management faculty practices: a Canadian perspective. Proceedings of the IATUL conference. Paper 2. Viewed 21 March 2017, from: *http:// docs.lib.purdue.edu/iatul/2015/mvd/2*

Smith, M. L. 2014. Data governance: where technology and policy collide. In Ray, J.M. (ed). *Research data management: practical strategies for Information professionals*. Indiana, US: Purdue University.

Social Sciences and Humanities Research Council (SSHRC). 2015. Tri-agencies adopt statement of principles on digital data management. Viewed 23 March 2017, from: *http:// www.sshrc-crsh.gc.ca*

The Society of College, National and University Libraries (SCONUL). 2015. Research data management: briefing for library directors. Viewed 24 July 2016, from: *http:// www.sconul.ac.uk/sites/default/files/documents/SCONUL%20RDM%20briefing.pdf*

Taylor, R. 2014. The SKA and the rise of big data in radio astronomy. Presentation given at the research Africa conference 2014, Cape Town, South Africa. Viewed 26 September 2016, from: http//www.eresearch.ac.za/presentations-2014-day-three

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Read, E. et al. 2011. Data sharing by scientists: practices and perceptions. PLoS One 6(6). Viewed 20 November 2016, from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101

Tibbo, H.R. 2003. On the nature and importance of archiving in the digital age. In M. Zelkowitz (ed.). *Advances in computers: Vol. 57. Information Repositories.* Retrieved June 23, 2016 from doi: 10.1016/S0065-2458(03)57001-2.

Thanos, C. 2010. Global research data infrastructures: the GRDI 2010 vision. Viewed 12 September 2016, from: *http://www.grdi2020.eu/Repository/FileScaricati/fc14b1f7-b8a3-41f8-9e1e-fd803d28ba76.pdf*

UK National Archives. 2012. Appraisal policy. Viewed 23 June 2016, from: *http:// www.nationalarchives.gov.uk/documents/information-management/ appraisal_policy.pdf*

United Nations Environmental Programme (UNEP). 2015. *Green economy sector study on agriculture in Kenya.* Nairobi-Kenya: UNEP. Viewed 3 October 2016, from*: http:// drustage.unep.org/greeneconomy/sites/unep.org.greeneconomy/files/publications/ kenya_agriculture_final.pdf*

University of California, Los Angeles. n.d. Data mining: what is data mining? Viewed 9 March 2017, from: *http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/ palace/ datamining.htm*

University of Cape Town (UCT). 2016. UCT libraries joins network of data and information curation communities. Viewed 4 April 2017, from: *http://www.eresearch.uct.ac.za/news/uct-libraries-joins-network-data-information-curation-communities*

The University of Western Australia. 2016. Metadata standards. Viewed 3 August 2016, from: *http://guides.is.uwa.edu.au/c.php?g=325196andp=2178564*

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. and Horton, L. 2011. *Managing and sharing data best practices for researchers.* London, UK: Data Archives.

Van Deventer, M. V. and Pienaar, H. 2015. Research data management in a developing country: a personal journey. *International Journal of Digital Curation*, 10(2), 33-47.

Van Wyk, J. and Van der Walt, I. 2014. Going full circle: research data management @ University of Pretoria. Presentation at eReseach Africa 2014 conference, held at University of Cape Town, Cape Town, South Africa 23-27 November 2014. Viewed 28 March 2016, from: *http://www.eresearch.ac.za/sites/default/files/image _tool/images/140/ Going_Full-Circle_RDM_UP_VanWyk_VanderWalt_26Nov_2014.pdf*

Whitehead, M. and Bourne-Tyson, D. n.d. Multi-stakeholder engagement in research data management. Viewed 30 March 2017, from: *https://webcache.googleusercontent.com/ search? q=cache:UxfMleJLe8MJ:https://portagenetwork.ca/wp-content/uploads/2016/ 06/ IATUL2016Multi_Stakeholder_Engagement_in_RDM.pdf+andcd=1andhl=enandct=clnk*

Whyte, A. and Allard, S. 2014. How to discover research data management service requirements. DCC how-to guides. Edinburgh: Digital Curation Centre. Viewed 7 July 2016, from: *http://www.dcc.ac.uk/how-*

Whyte, A. and Wilson, A. 2010. How to appraise and select research data for curation. Edinburgh: Digital Curation Centre. Viewed 23 June 2016, from: *http://www.dcc.ac.uk/ resources/how-guides/appraise-select-data*

Witt, M. 2008. Institutional repositories and research data curation in a distributed environment. *Library Trends*, 57(2), 191-201.

The World Agroforestry Centre. 2012. Research data management training manual. Viewed 29 March 2016, from: *http://worldagroforestry.org/research-methods/sites/default/files/ Research% 20Data%20Management%20Manual%20-%20web%20temp.pdf*

Yiotis, K. 2008. Electronic theses and dissertation (ETD) repositories: what are they? Where do they come from? How do they work? *OCLC Systems and Services: International Digital Library Perspectives*, 24, 101-115.