

Submitted on:07.07.2017

Research data management services: are academic libraries in Zimbabwe ready? The case of University of Zimbabwe library

Nancy Nhendodzashe
University of Zimbabwe library
University of Zimbabwe
630 Churchill avenue, Mount Pleasant, Harare, Zimbabwe
263-4-303211
nnhendodzashe@uzlib.uz.ac.zw

Notice Pasipamire
Department Library and Information Science
National University of Science and Technology
P.O. Box AC 939, Ascot, Bulawayo, Zimbabwe
263-9-282842
notice.pasipamire@nust.ac.zw



Copyright © 2017 by Nancy Nhendodzashe and Notice Pasipamire. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

This research paper reports on the readiness of University of Zimbabwe (UZ) library to offer research data management (RDM) services after the library indicated that they are planning to offer RDM services to capture the ever-increasing research data from the institution and to help researchers meet the funders' mandates for RDM. The investigation used the Technological, Economic, Legal, Organisational and Scheduling (TELOS) feasibility model and the research data life cycle as means to assess the readiness of UZ using both qualitative and quantitative methods. Questionnaires were used to gather information from researchers while interviews were used to collect data from faculty librarians, university librarian and the library information and communication technology (ICT) manager. Stratified and judgmental sampling methods were used to select participants for inclusion in the study. The study findings revealed that the UZ library was partly prepared as it had the necessary technological infrastructure and economic resources but did not have the enabling legal framework and skills to provide RDM services throughout the data lifecycle.

Keywords: Research data management, University of Zimbabwe Library, Feasibility, Research support.

1.0 Introduction

Research data management services are being established in response to the changing scientific research landscape and are being supported both by the existence of cyber infrastructures and data sharing mandates by funding agencies and researchers who are committed to open science, which advocates for open verification and reproduction of

research data (Naum, 2014). To this end, researchers are turning to libraries for research data management services as many academic institutions, research funders and journal publishers are beginning to request research data as part of the audit trail and open data policy. Offering research data management services is an emerging trend in most academic institutions in developing countries such as Zimbabwe and like any other intervention, organisations are required to investigate the needs of researchers, the readiness and capabilities of on-campus partners, and available infrastructure on a particular campus prior to introducing RDM services. Lotter (2014) encouraged that before setting up any RDM services; the infrastructure and investment need to be carefully assessed by each institution in line with their own mission, objectives and strategic aims.

1.2 Literature Review

A cursory review of literature shows that very few studies were conducted to assess the feasibility of offering RDM. Sykes (2009) assess the feasibility of a national shared service for managing research data in the United Kingdom focusing on technological requirements for the entire data lifecycle. The study revealed that there was an acknowledged difficulty for researchers in retaining or managing research data beyond the life of a project once the funding associated with the project ceases. Der (2015) explored the academic libraries' readiness for research data management from Hungary and Estonia academic libraries which revealed that libraries in these areas were indeed at the beginning of addressing the issue, as at the time of the investigation, only a small amount of the participating institutions had services in place for supporting research data management at their institutions. In South Africa Chiware and Mathe (2015) conducted a study to establish the requirements for setting up RDM services at the Cape Peninsula University of Technology (CPUT) Library. It was found that there was a great need for structured RDM services and tools for setting up RDM platforms that include technology, staff and policies within the institution. From literature accessed only one study was carried in Zimbabwe by Ndhlovu (2016) who looked at the preparedness for digital curation and preservation at the National University of Science and Technology. He concluded that there were inconsistencies in digital curation and ICT competences were low among some library staff. No studies were conducted at the UZ on the feasibility of offering RDM services.

Pinfield et al. (2014) highlighted that RDM is underpinned by processes with technical implementations including data repositories, hardware and software allowing for storage and transport of different types and quantities of data. Berman (2008) proposes that to ensure good data stewardship, institutions must be aware of data costs and include them in their overall information technology (IT) budget. Hole et al. (2010) assert that predicting the costs of long-term digital collection, storage, preservation of and access to research data is a crucial yet complex task for even the largest repositories and institutions. The costs involved include costs for staff recruitment and training, consumables, equipment and maintenance. Davidson et al. (2014) posit that preserving research data for the long term has a cost; although the infrastructure itself is costly, more significant is the cost associated with human resources, such as personnel to manage and maintain the archive. Storage costs for digital data are decreasing, but costs related to storage, such as power, data curation and annotation, and personnel, are not decreasing (Berman 2008 as cited by Strasser, 2014).

Greenbaum and Gerstein (2003) acknowledged that it has become increasingly apparent that to achieve seamless access to data it is necessary not only to adopt appropriate technical standards, practices and architecture, but also to develop legal frameworks that facilitate access to and use of research data, whether on an inter-organisational basis or across national borders. The UK Data Archive (2015) emphasises that before embarking on a RDM project, it is imperative to know your legal, ethical and other obligations regarding research data, towards research participants, colleagues, research funders and institutions.

It is crucial that researchers ensure ownership of primary materials and research data is identified and documented at the start of a research project and reviewed and updated whenever appropriate (Strasser, 2015; Charles Darwin University, 2010; Fitzgerald and Pappalardo, 2007). Fitzgerald and Pappalardo (2007) suggest that reference to the “owner” of data will usually be a reference to the person who owns legal rights in relation to the data, the person who has physical possession of the data, the person who controls access to and use of the data or a combination of any of the above.

The role of RDM policies cannot be overemphasised in the success of RDM programmes (Higman and Pinfield, 2015; Nugroho et al., 2015). RDM policies respond to a number of drivers, including data collection, storage and preservation, but also data access and sharing (Pinfield et al., 2014). Policies are mostly used to provide “credentials” for those

championing RDM, gain access to funding for IT infrastructure (Pryor, Jones and Whyte, 2014), clarify institutional positions and outline roles and responsibilities (Brown and White, 2014).

Henderson and Knott (2014) observed that the introduction and success of RDM services in academic libraries calls for the need to hire new staff or re-skilling and up skilling of librarians to take up new roles and responsibilities. As the creators and users of research data, researcher engagement is crucial in the design and development of RDM services.

1.3 Conceptual Framework

The study combined the Technological, Economic, Legal, Organisational and Schedule (TELOS) feasibility framework with the UK Data Archive Research Data Lifecycle model to assess institutional readiness to offer RDM services by examining the enabling factors for setting up RDM services. A research data lifecycle describes the stages through which research data is collected, recorded, processed and results are published (UK Data Archive, 2015). The UK Data Archive Research Data Lifecycle provides a data lifecycle model as an aid to researchers considering how data management relates to the lifecycle of a research project (UK Data Archive, 2015). Figure 1.1 below shows the resultant conceptual framework which was developed from the two models.

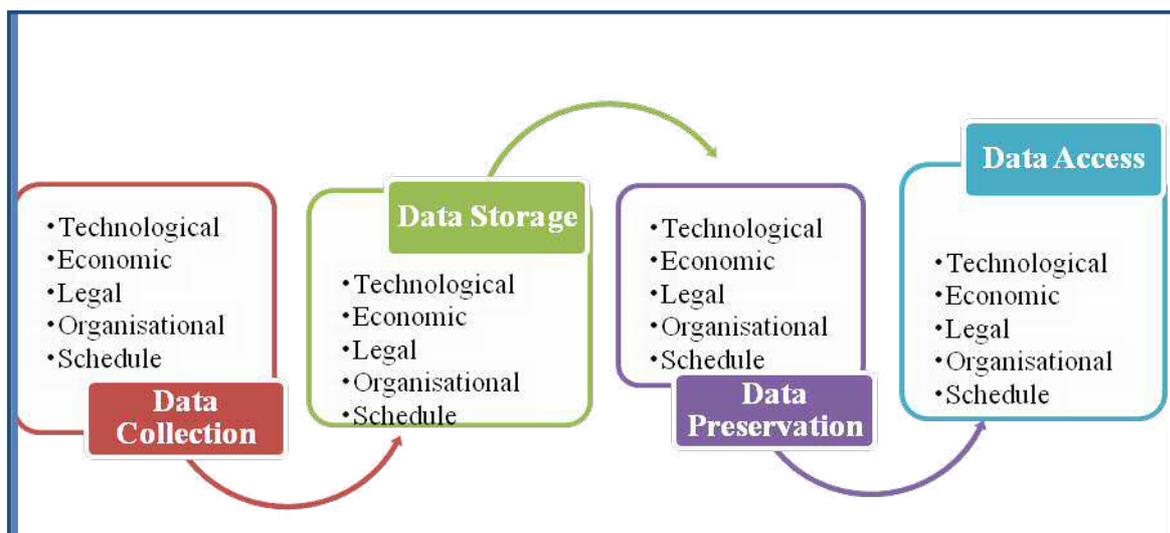


Figure 1.1 Conceptual Framework: Application of the TELOS framework to the revised UK Data Archive Research Data Lifecycle (Researchers)

The first stage from the conceptual framework of the research data lifecycle model is data collection. This involves gathering research data from researchers using various media. At this stage, research libraries should be in a position to offer advice on the appropriate structure, storage of research data (Stuart, 2014). Data storage is the next stage in which libraries can provide RDM services. After data has been acquired from different researchers, there has to be mechanisms in place to store the data. The University of Queensland (2017) highlights that librarians must ensure that all research data, regardless of format, is stored securely and backed up or copied regularly

Data preservation is the next stage and it involves migrating data to suitable formats and media for preservation, creating backups, and creating any additional metadata that is necessary for preservation (Strasser, 2014). Data access is whereby an institution distributes data, share data, control access, establish copyright and promote data (UK Research Data Archive, 2015). Providing access to the data not only requires that the data is available, but that it can be found and appropriate rights are provided for its reuse.

1.4 Methodology

Both quantitative and qualitative methods were adopted and data was collected from researchers (138), faculty librarians (9) university librarian (1) and library ICT manager (1) using questionnaires and interview techniques. Stratified random sampling was used to select researchers while participants from the library were selected using judgemental sampling. From the questionnaires which were distributed to researchers, 104 were returned with complete information. The questionnaire gathered information about the types and quantities of research data generated by researchers and to find out their attitudes towards RDM. Interviews were conducted with faculty librarians to establish their skills, knowledge and attitudes towards RDM. Other interviews were held with the university librarian to find out the administrative requirements for establishing RDM services and the library ICT manager to find out the technological requirements for setting up RDM services.

1.5 Findings and Discussion

Findings of the study are presented according to the conceptual framework developed for the study consisting of technology, economic, legal, organisational and scheduling as variables measured around the research data life cycle.

1.5.1 Information Technology Requirements for RDM

In terms of technological requirements for data collection, storage, preservation and access the study found that the library made use of distributed servers and some projects share servers that have a storage capacity of one Terabyte to store information for the repositories and other information that is relevant to the library. The ICT department pointed that if the library is to start RDM services, the current servers can only store data but when it comes to preservation and data backups; there is need for an upgrade of the current servers because UZ researchers produce different types of research data. Most researchers reported that they generate data in text documents (87%) followed by spreadsheets (62%). The findings are consistent with those of Van Tuyl and Michalek (2015) who found that the most common data types at the Carnegie Mellon University were text documents and spreadsheets. The data types may be considered low-difficulty formats for preservation purposes since there are clear transformations or migration pathways for these formats and the UZ libraries have extensive experience working with these types of documents. It was found that a few researchers (4%) and (3%) generate software applications source code and configuration data respectively. These two categories might pose more difficult to preserve, especially when considering the potential for obsolescence of formats and software applications for using these files and the need for an emulation strategy to preserve a functional environment to run them in the future (Van Tuyl and Michalek, 2015).

1.5.2 Economic Needs for RDM

The findings revealed that acquiring, maintaining and upgrading ICT resources and staff training would be the major cost for the project. The head of the library noted that the costs mainly revolve around the infrastructure needed especially for storage and preservation of data sets. It was further noted that, “Staff training is another significant cost that we can consider given that RDM is an emerging field that has not been incorporated into the current curricula for library and information science.” Other costs include preparing data and metadata, and advocacy costs. The library ICT department noted that if the library was going

to use existing repository systems, costs will be incurred in acquiring additional servers for data storage and preservation and also training on RDM. Other costs that were highlighted include the costs of maintaining and upgrading the data repository once it is set. It is interesting to note that the UZ library had mechanisms in place to cater for the anticipated costs for setting up RDM.

1.5.3 Legal Obligations for RDM

The study findings revealed that the legal requirements for the collection, storage, preservation and access to research data include copyright issues, institutional RDM policy, access control, and data protection. Researchers were asked if they had any legal concerns regarding the use of their research data. The results revealed that 51% had legal concerns while 49% did not have any concerns. For those who said they had legal concerns, most researchers (89%) cited intellectual property rights issues particularly copyright and 11% mentioned data protection issues. It was found that the University of Zimbabwe did not have any policies that govern RDM activities of researchers but had an Intellectual Property Rights (IPR) policy which caters for copyright issues. As a result, the library had intentions to develop a RDM policy at the start of the implementation of RDM services. This is in accordance with the recommendations given by Patel (2016) who recommends that before embarking on the RDM journey there has to be an institutional RDM policy that clearly spell out the following in the context of data management: purpose, scope, applicability and guidelines to the data contributors relating to data submission, licensing, metadata entry, data classification, copyright agreements and conditions under which the data withdrawal requests, if any, will be considered, terms and conditions of the use of data, protection of confidentiality of sensitive data, protection of data against security breaches and intellectual property concerns.

1.5.4 Organisational Capabilities for RDM

The results of the study showed that librarians need to possess the following skills in order to take up RDM roles: ICT skills, digital curation, advocacy and training skills. Most faculty librarians reported that there was need for data curation skills (7), ICT skills (9), marketing and advocacy skills (6), scholarly communication (3), knowledge of the research process (4), metadata description (9), RDM workflows and systems (5), data recruitment (2) and

understanding of the different types of research (3). Similarly, Cox and Pinfield (2014) found that the specific skills needed for RDM include data curation skills, technical IT skills and knowledge of research methods and the need for disciplinary knowledge. However, the findings of the study revealed that currently the UZ library staff did not possess the requisite skills and knowledge for research data collection, storage, preservation and access. The findings fit the researcher's anecdotal understanding that none of the faculty librarians had any course in RDM during their training as the area was not covered in the past Library and Information Science curriculum. The result shows that librarians need to be trained on all the activities that take place throughout the research data lifecycle.

1.5.5 Management Support to RDM Services

It was established that UZ top management always support library projects in terms of infrastructure, finances and human resources. Regarding the proposed RDM programme, some librarians said that management support is needed in securing funds for staff training, acquisition of ICT resources for data storage and preservation and in enacting and ratifying enabling policies. However, three (3) faculty librarians reported that when it comes to library policies, there is need for improvement. They cited an example of the Open Access Policy which was submitted in 2012 for adoption but until now it was not approved and there are no mandates despite concerted effort from the library. One faculty librarian emphasised that management support should be all encompassing if RDM is to succeed at the UZ.

1.5.6 Scheduling RDM services at the UZ

Researchers were asked about how soon they require RDM services. The study revealed that the majority of the researchers (77) require RDM services in the next one year, 9 said they require RDM services in 2-3 years' time, 3 said in 3-5 years and 3 were not sure. Figure 1.2 below shows the responses which were given.

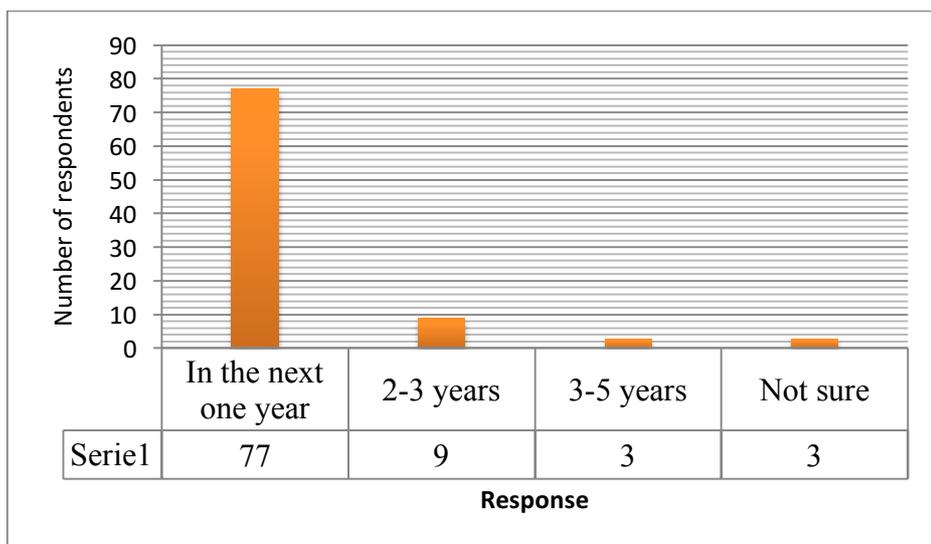


Figure 1.2 Expected time for RDM services

Figure shows 1.2 shows that most researchers require RDM in the next one year. From the library side, it was revealed that they plan to offer RDM services in 2017. The library ICT manager was of the view that it “takes a few days if all the necessary resources are in place”. On the contrary, Steeleworthy (2014) argues that developing an IT and staffing strategy to provide the technology and support that maintain an RDM programme is not always feasible in the short-to-medium term.

The study findings also revealed that most researchers prefer a centralised multidisciplinary data repository. The data shows that over half of the respondents 70% (73) preferred a centralised multidisciplinary data repository. Ten percent (10%) (10) preferred a centralised discipline specific data repository, 7% (7) preferred a decentralised discipline specific data repository while 5% (5) preferred a decentralised multidisciplinary data repository. The remaining 9% (9) said they did not know how RDM services could be implemented. Likewise, faculty librarians and the university librarian all agreed that there is need to provide a service that caters for all academic disciplines so that other faculties will not feel left out. On the contrary, previous studies provided that there are factors such as the availability resources, skills and technology that have a bearing on whether an institution should take a campus wide approach or departmental approach to RDM (Shen and Varvel, 2013). There are mixed views from literature on how RDM services can be rolled out with one of the main themes being to start with a small number of RDM services, and then look to expand those services (Toups and Hughes, 2013; Wright et al., 2014). However, research has shown that this approach may prove to be problematic if the demand for the service is high.

1.6 Conclusion

The study established that RDM is a service that researchers require from the library; however, the UZ library is partly ready to offer this service. In terms of the technological requirements the current setup can handle research data, and it appears management is willing to provide financial support to ensure that the project is a success. However, the UZ library is not prepared in terms of the RDM legal obligations and skills that are required to take up RDM activities. Given that researchers are worried about copyright issues of their data and how data are going to be accessed, it is imperative that the library enact enabling policies for access, control, data protection as well ownership of data. Staff development is another important issue that needs attention if the project is to be a success.

References

Berman, F. 2008. Got data?: a guide to data preservation in the information age. *Communications of the ACM* 51, 50.

Charles Darwin University. 2010. Research Data Management Procedures. Available at: <http://www.cdu.edu.au/governance/doclibrary/pro-123.pdf> (Accessed 2017 April 28)

Chiwere, E., and Mathe, Z. 2015. Academic libraries' role in Research Data Management Services: a South African perspective. *South Afr. J. Libr. Inf. Sci.* 81, 1. Available at: <http://sajlis.journals.ac.za/pub/article/view/1563> (Accessed 2017 January 15)

Der, A. 2015. Exploring the academic libraries' readiness for research data management: cases from Hungary and Estonia. Available at: <https://oda.hio.no/jspui/bitstream/10642/3367/2/Der.pdf> (Accessed 2017 January 12)

Fitzgerald, A. and Pappalardo, K. 2007. Building the infrastructure for data access and reuse in collaborative research: An Analysis of the Legal Context. Available at: <http://eprints.qut.edu.au/8865/1/8865.pdf> (Accessed 2017 February 11)

Greenbaum, D. and Gerstein, M. 2003. A universal legal framework as a prerequisite for database interoperability. *Nature Biotechnology* (21) 979.

Henderson, M.E., and Knott, T.L. 2015. Starting a Research Data Management Program Based in a University Library. *Med. Ref. Serv. Q.* 34, 47–59.

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 systems*, 49(4), pp.364-381. doi: 10.1108/PROG-01-2015-0005

Lotter, Lucia. 2014. “Reflections on the RDM Position in South Africa.” Paper presented at the LIASA Research Data Management Workshop, Cape Town, South Africa. Available from http://www.dcc.ac.uk/webfm_send/1631 (Accessed 2016 December 13)

Naum, A. 2014. Research data storage and management: library staff participation in showcasing research data at the University of Adelaide. *Aust. Libr. J.* 63, 35–44.

Ndhlovu, P. 2016. The state of preparedness for digital curation and preservation: a case study of the National University of Science and Technology Library. Bulawayo: National University of Science and Technology

Nugroho, R.P., Zuiderwijk, A., Janssen, M. and de Jong, M. 2015. A comparison of national open data policies: lessons learned. *Transforming Government: People, Process and Policy*, 9(3), pp.286-308. Available at: <http://dx.doi.org/10.1108/TG-03-2014-0008>

Patel, D. 2016. "Research data management: a conceptual framework ", *Library Review*, Vol. 65 Iss 4/5 pp. 226 - 241 Available at: <http://dx.doi.org/10.1108/LR-01-2016-0001> (Accessed 2017 January 25)

Pinfield, S., Cox, A.M., and Smith, J. 2014. Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLOS ONE* 9, e114734. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114734> (Accessed 2017 January 6)

Shen, Y., and Varvel, V.E.J.2013. Developing Data Management Services at the Johns Hopkins University. *Journal of Academic Librarianship* 39, no. 6 Pages Doi: 552–557. doi:10.1016/j.acalib.2013.06.002

Soehner, C., Steeves, C., and Ward, J. 2010. E-Science and Data Support Services: A Study of ARL Member Institutions. *Assoc. Res. Libr.*

Strasser, CA. 2014. *Data Management for Libraries: A LITA Guide*, ALA TechSource, New York, US. Available from: Proquestebrary. (Accessed 15 December 2016).

Sykes, J. 2009. Managing the UK's Research Data: Towards a UK Research Data Service. *New Review of Information Networking*. 14(1), 21-36. Available at: <http://www.tandfonline.com/doi/full/10.1080/13614570902953556> (Accessed 2016 November 18)

Tenopir, C., Sandusky, R.J., Allard, S., and Birch, B. 2013. Academic librarians and research data services: preparation and attitudes. *IFLA J.* 39, 70–78.

Tenopir, C., Sandusky, R.J., Allard, S., and Birch, B. 2014. Research data management services in academic research libraries and perceptions of librarians. *Libr. Inf. Sci. Res.* 36, 84–90. Available at: <http://www.sciencedirect.com/science/article/pii/S0740818814000255> (Accessed 2017 January 16)

UK Data Archive. 2015. Create and manage data: Research Data Lifecycle. Available at: <http://www.data-archive.ac.uk/create-manage/life-cycle> (Accessed 2016 October 5)

Van Tuyl, S. and Michalek, G. 2015. Assessing Research Data Management Practices of Faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication*, 3(3), eP1258. Available at: <http://dx.doi.org/10.7710/2162-3309.1258>