Research data management and services: Resources for different data practitioners

Gilbert Exaud Mushi[1]

Abstract

The emergence of data-driven research and demands for Research Data Management (RDM) has created interest in global academic institutions and research organisations. Some of the libraries, especially in developed countries, have started offering RDM services to their communities. Although lagging, some academic libraries in developing countries are planning or implementing the service. However, the level of RDM awareness is deficient among researchers, librarians and other data practitioners.
This paper aims to present available open resources for different data practitioners, particularly researchers and librarians. It includes training resources for researchers and librarians, Data Management Plan (DMP) tool for researchers, a data repository available for researchers to freely archive and shares their research data to the local and international communities.

A case study with a survey was conducted at the University of Dodoma to identify relevant RDM services so that librarians could assist researchers in making their data accessible to the local and international community.

The study findings revealed a low level of RDM awareness among researchers and librarians. Over 50% of the respondent indicated their perceived knowledge as poor in the following RDM knowledge areas; DMP, data repository, long term digital preservation, funders RDM mandates, metadata standards describing data and general awareness of RDM. Therefore, this paper presents available open resources for different data practitioners to improve RDM knowledge and boost the confidence of academic and research libraries in establishing the service.

1. Introduction

Sustainable development and economies in the fourth industrial revolution are purely characterised driven by data-intensive research and innovation. The massive data generated through research, social-economic activities, and advanced technologies trigger the need to capture, integrate, and interpret new competitive knowledge. Governments and federal granting agencies, especially in developed countries such as the National Science Foundation (NSF), National Institute of Health (NIH) in the US, Australian Research Data Commons and e-Science Core programme in the UK, are championing this movement by mandating RDM practices in research organisations and academic institutions (Chiware and Mathe, 2016; Tang and Hu, 2019). All research organisations using public funds must manage their research data so that valuable data can be stored, shared and accessed for the long term. Researchers must submit data management plans (DMPs) alongside their proposals as a mandate from their organisations and research funding agencies.

Many research organisations and academic institutions in the US, UK, Australia and other developed countries have somewhat experience offering data management services to their research community and internationally. National Research Foundation (NRF) in South Africa mandates researchers and research institutions to effectively manage data during research and share valuable data to national and

international communities (Mushi, 2017b)**.** Having realised the potential benefits of data management and sharing, some of the research institutions established the RDM initiative without the mandates of the federal funding agencies or governments.

Data literacy plays a significant role in the growth of research and innovation, sustainable economic development and social well-being of society. Thus, attracting local and international collaboration projects to solve global challenges such as food security and climate change.

2. Statement problem

RDM has been a subject of interest for over a decade, with several academic and research libraries providing and planning for the project. The tremendous growth of academic libraries offering RDM services has been reported, especially in the developed countries (Cox *et al.*, 2017). Although lagging, research and educational institutions in developing nations plan to be part of this changing research environment. A case study conducted at the University of Dodoma, Tanzania, on identifying relevant RDM services for the library depicts a low awareness of RDM by researchers and the academic community. However, most researchers and postgraduate students were willing to practice RDM and share their research data with the local and international community. Therefore, this paper presents available open resources for data practitioners, particularly researchers and librarians. These include training materials and online courses, DMPs tools, and open data repositories for researchers to archive their valuable data freely.

3. Literature review

The new data-intensive and technology-based research environment jointly require different stakeholders' responsibilities, unlike a traditional research environment whereby in most cases, a researcher or group of researchers solely conduct research and manage data through its life cycle. It involves researchers, librarians, data analysts, ICT personnel and more – it depends on the nature of research and organisation culture. Researchers and librarians need to acquire skills (data literacy) and knowledge to navigate this new paradigm effectively. Researchers need to present a data blueprint showing how data will be collected, stored, preserved and shared to meet research funders requirements (Tenopir, Birch and Allard, 2012; Tang and Hu, 2019).

Librarians and researchers work together with librarians' roles and responsibilities revolving around the research data life cycle. There are different librarian roles in each stage of the data life cycle which begins with planning. Planning requires researchers to show a data roadmap such as how data will be collected, secured, stored, shared and reused. It involves the use of so-called data management plans (DMPs). Data collection follows after a project that requires librarians to ensure data are collected based on respective DMPs using correct delimiters and consistent codes in data collection. Furthermore, librarians assure the quality of the collected data and describe data by assigning proper metadata to enable data finding, understanding and re-usability. They are also responsible for data preservation and deposit to a trusted data repository while ensuring the use of the friendly format and support data discovery and reuse (Tenopir, Birch and Allard, 2012; DataONE, 2020).

Librarians have for a long time served researchers, especially with literature and information literacy skills. Thus, making it a possible and favourite group to support researchers with RDM services. However, the most common setback for the librarians in this new role is the lack of RDM skills and knowledge (Chiware and Mathe, 2016; Yoon and Schultz, 2017). A study conducted in Tanzania also observed low awareness

of the RDM practices and associated benefits (Mushi, 2017). Mushi recommended advocacy of the services and skills needed to researchers and librarians through seminars, workshops and training.

4. Methodology

A case study used a purposive sampling technique to interview two (2) university managers (Director of Research and Publications and Director of Library Services). The study aimed to gather information on the managers' perception and future of RDM services in academic libraries. Using the snowball sampling technique, a researcher collected data from six (6) researchers and six (6) postgraduate students who could have experience with international funders' mandates when applying for research grants. Data were collected from 12 respondents from the University of Dodoma using Google forms online survey. The majority of respondents had never had encountered any RDM practice mandate because of the use of public funds from the government channelled to the university. Some other researchers were self-financing their research projects (Mushi, 2017a).

5. Results and discussion

5.1 RDM awareness and skills

The study observed the low awareness of RDM practices among researchers, with more than 88% of researchers storing and handling research data in their devices. They also rated their RDM skills and knowledge as poor. They suggested training interventions in different RDM skills areas, including a guide on using DMPs, a guideline for data appraisal, trusted data repositories, and other RDM practices presented in Figure 1 below.
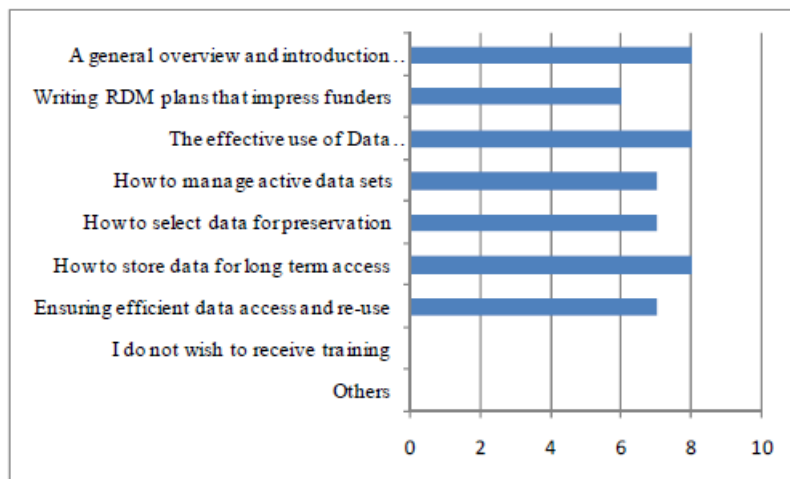

Figure 1. Training asked for by the researchers (Source: Mushi, 2017b)

5.2 RDM implementation model

Building from Jones et al. (2013), the study developed an RDM implementation model for implementing RDM services in academic libraries. This model has four implementation phases whereby phase one is on 'strategy, policy, procedures and infrastructure'. The phrase stands as a master plan for sustainable and effective RDM services at the institution. It involves developing a policy that specifies the responsibilities of different stakeholders in the organisation, developing RDM guidelines and necessary infrastructure. Phase two is about 'awareness creation, skill development and repository content'. Phase three insists on 'management of active data' while phase four is about 'data selection and preservation. Librarians, policymakers, and researchers can read about this four-phase RDM implementation model in the article

entitled '*Identifying relevant research data management services for the library at the University of Dodoma, Tanzania'* (Mushi, Pienaar and Deventer, 2020).

5.3 Identified RDM resources for different data practitioners
The study objective also aimed to identify different RDM resources available that can be used by various data practitioners when managing data through its life cycle. It includes training resources, online tools and forums that researchers can use, librarians, ICT personnel supporting RDM, senior managers, research funding agencies and other research stakeholders as described below.

5.3.1 RDM training materials and courses
Several institutions and organisations offer training materials to improve data literacy among data practitioners, especially researchers and librarians. The materials address the common challenge of lack of skills and knowledge to manage the new role of this exciting subject. Macdonald and Rice (2013) argued that librarians need to acquire new skills to build capacity for supporting RDM services at the institutions. These include the following:

MANTRA free online course. It provides online materials and training on RDM skills targeting graduate, postgraduate students, early researchers and professionals from different research disciplines. It offers training on various modules covering data services through its life cycle, including DMP, organising data, files format and transformation, documentation and metadata citation, storage and security, data preservation, rights and access, sharing protection and licensing. Edinburgh University of the UK maintains this online training programme. Training materials are licensed under the creative commons attribute, which means they can be used and shared while constantly acknowledging the source.
Access: https://mantra.edina.ac.uk/

DCC training and reference materials. Free online training offered by the UK experienced trainers in-depth subject experts in digital curation and research data management. It has two training approaches; one is an inclusive training event for all, and the second is customised private training requested by the individual or an organisation. DCC can tailor the latter approach to meet the requirements of the requesting individual or organisation. It targets different data practitioners and research stakeholders from institutional managers, researchers, students, professionals, librarians and other research support staff. Digital Curation Center (DCC) of the UK maintains the training resources.
Access: http://www.dcc.ac.uk/training

University of Minnesota Libraries Consultation and Workshop. Apart from providing data management consulting services to its community, it offers learning materials in various formats such as templates, slides, and videos, which improves skills in general data management, developing a DMP, data sharing, access and ownership, and organising and storing data. Resources are helpful to all research staff dealing with data and supportive services.
Access: https://www.lib.umn.edu/datamanagement/workshops

Supporting Data Management Infrastructure for Humanities (SUDAMIH). SUDAMIH is an RDM support infrastructure for the humanities helping institutions to manage research data effectively addressing technical requirements. It facilitates the creation of online databases to collect data in humanities and support different data services such as text, image, and geo-data. Although materials are meant to be used in UK based institutions, they are also valuable for other institutions wishing to implement RDM services in humanities. The project is maintained by the Oxford University of the UK and funded by the Joint Information System Committee (JISC).

Mushi, Gilbert Exaud (2021), Research data management and services: Resources for different data practitioners, IASSIST Quarterly 45(3-4), pp. 1-8. DOI: https://doi.org/10.29173/iq995

Access: http://sudamih.oucs.ox.ac.uk/index.xml

RDMRose. The training project was for information professionals from UK leading iSchool targeting LIS educators and provide continued professional development (CPD) learning materials in RDM. Although this project was closed, training materials remained available and valuable for information professionals planning RDM services.
Access: https://www.slideshare.net/rdmrose

Essentials 4 Data Support. Provide support and training courses on data storing, managing, archiving and sharing. The site is maintained by Research Data Netherlands (rdnl) and offers various courses suitable for librarians, archivists, researchers, managers and related research stakeholders in RDM services.
Access: https://researchdata.nl/

Digital Humanities Data Curation (DHDC). Provides different resources, including articles and RDM guides which are helpful to researchers and librarians in humanities. Resources provided in this platform are published under the Creative Commons No-Commercial Attribution license to encourage broad access and reuse of globally educational resources.
Access: https://guide.dhcuration.org/about/

Coursera – Research Data Management and Sharing. Provides opportunities for data practitioners such as librarians and researchers to join their platform and access training courses and materials freely. Financial aids are available for some individuals who could meet the criterion provided for the course participants. The University North Caroline jointly offers the courses at Chapel Hill and the University of Edinburgh.
Access: https://www.coursera.org/learn/data-management

New England Collaborative Data Management Curriculum (NECDMC). It is an educational and RDM training resource that teaches RDM best practices explicitly to students and researchers in health sciences and engineering. Although the teaching curriculum is designed based on NSF data plans recommendations, the materials are still valid and valuable to other institutions around the globe as it addresses many common challenges of RDM services. Materials are published under Creative Commons Attribution-Non-Commercial Share-Alike, giving freedom to reuse, share, and modify while acknowledging the source. The platform is maintained by Lamar Soutter Library of Massachusetts Medical School in partnership with other libraries in the New England region.
Access: https://library.umassmed.edu/resources/necdmc/index

CESSDA- free online expert tour guide to data management. Offers training materials that aim at imparting skills in managing data in all stages of the research data life cycle. Materials are organised in different formats, such as texts and videos. Furthermore, the platform offers a guide to researchers on managing data and making their data FAIR (findable, Accessible, Interoperable and Reusable). It is designed and maintained by European experts and is mainly on social sciences research datasets.
Access: https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide

UK Data Archive. Experienced platform offering data management resources, including training and publications of all matters related to research data. It provides opportunities for librarians and researchers to gain and improve their skills in RDM services. The long experience of training in RDM led to the

handbook published by SAGE publication Ltd. In 2014 which is entitled *"Managing and Sharing Research Data: a guide to good practices".*
Access: https://www.ukdataservice.ac.uk/manage-data/training.aspx

DataOne training resources. It is a freely available education resource for librarians and researchers on RDM. A comprehensive package covering all significant aspects of data management, general under the Creative Commons Zero license (CC0), gives users enormous freedom to use, share, and edit even without acknowledging the sources. Resources that are accessible include sample assignments used for seminars or workshops, slides and handouts are covering essential RDM services, DMP, policy issues, metadata and data protection. The US National Science Foundation funds the platform (NSF).
Access: https://www.dataone.org/education-modules

MIT Libraries RDM workshops. MIT libraries offer many different workshops to build capacity in managing research data. The topics cover general RDM overview, DMP, data storage and sharing, file organisation and version control. The target group of these workshops includes postgraduate and graduate students, academic staff, researchers, librarians, and other teams working with data to improve their skills to manage their new roles.
Access: https://libraries.mit.edu/data-management/services/workshops/

5.3.2 DMP tools
The data management plan is an essential tool required by most research funders nowadays. It provides a roadmap of data through its life cycle while meeting standards set by the institution of funding organisation. Librarians should have the skills to assist researchers in creating an impressive DMP for their research projects. Most institutions already offering RDM services have developed their own DMP tool reflecting their requirements and that of their federal government agencies. However, researchers and institutions willing to practice data management can use DMP tools open for use by the research community globally. These tools are not tailored to meet a specific institution of funder requirements but address common issues of DMP that are accepted by many funders and institutions. Some of these DMP tools have options for librarians and researchers to develop DMP based on particular funder requirements. Moreover, other DMP tools such as DMPonline maintained by Digital Curation Center allow institutions to customise the template to provide tailored support for their users (chargers apply for institutions in this service). Most DMP tools are free to use by individual researchers and librarians. Below is the summary of DMP tools librarians and researchers can use to develop a data management plan (see table 1 below).

Table 1. Data Management Plan tools

S/N	DMP Tool	Organisation	Accessed at:
1	Data Management Planning Tool (DMPTool)	University of California Digital Curation Centre of the California Library	https://dmptool.org/
2	DMPonline	Digital Curation Centre	http://www.dcc.ac.uk/dmponline
3	Perdue Data Curation Profile Toolkit	Perdue University Libraries	http://datacurationprofiles.org/

5.3.3 Data repositories

There are also various data repositories offering free data archiving services for researchers. It provides opportunities for librarians and researchers to deposit their research data in a manner that can be FAIR (Findable, Accessible, Interoperable, and Reusable). Librarians need to be aware of these resources to assist researchers in preparing data for archiving, especially assigning good metadata for easy discoverability of data. There are many discipline-specific public data repositories. However, this article focuses on generalist data repositories that researchers and institutions can archive research data (see Table 2 below).

Table 2. Generalist repositories

S/N	Data Repository Name	Fees/Cost	Accessed at:
1	Zonedo	Free (Donations are encouraged)	http://help.zenodo.org/
2	Mendeley Data	Free up to 10 GB datasets	https://data.mendeley.com/
3	Dryad Digital Repository	$120 USD for first 20 GB, and $50 USD for each additional 10 GB	http://datadryad.org/
4	Figshare	100 GB free per *Scientific Data* manuscript.	http://figshare.com/
5	Harvard Dataverse	Free	http://dataverse.harvard.edu/
6	Open Science Framework	Free	http://osf.io/

Source: (Spinger Nature, 2020)

6. Conclusion

All RDM resources discussed in this article provide opportunities for the libraries, researchers and other data champions to improve their data literacy and begin practising and offering RDM services in a limited resource environment of academic and research libraries. The resources address many data management challenges, particularly the required skills of researchers, librarians and research support staff. Developing RDM services in the institution requires many resources more than that of an institutional repository. Using DMP tools and data repositories services offered by other institutions encourages researchers and institutions, especially from developing countries, to be part of the changing research environment which is more data-intensive and collaborative. The discussed resources are essential to increase data literacy which will eventually stimulate data management practices for the sustainable development of societies.

7. Reference

Chiware, E. R. T. and Mathe, Z. (2016) 'Academic libraries' role in Research Data Management Services: a South African perspective', *South African Journal of Libraries and Information Science*, 81(2), pp. 1–10. DOI: https://doi.org/10.7553/81-2-1563.

Cox, A. M. *et al.* (2017) 'Developments in research data management in academic libraries: Towards an understanding of research data service maturity.', *Journal of the Association for Information Science and Technology*, 68(9), pp. 2182-2200. DOI: https://doi.org/10.1002/asi.23781.

DataONE (2020) *Best Practices, DataONE: Data Observation Network for Earth*. Available at: https://www.dataone.org/best-practices (Accessed: 23 April 2020).

Jones, S., Pryor, G. and Whyte, A. (2013) 'How to Develop Research Data Management Services - a guide for HEIs', *Digital Curation Centre*, (March), pp. 1–22. Available at: http://www.dcc.ac.uk/resources/how-guides.

Macdonald, S. and Rice, R. (2013) *'DIY' Research Data Management Training Kit for Librarians*, *DigCurV: Lifelong Learning Programme*. Available at: http://datalib.edina.ac.uk/mantra/libtraining.html.

Mushi, G. E. (2017a) 'Identifying and Implementing Relevant Research Data Management Services for the Library at the University of Dodoma, Tanzania [Data set].' Available at: https://zenodo.org/record/3553837.

Mushi, G. E. (2017b) *Identifying Relevant Research Data Management Services for the Library at University of Dodoma, Tanzania*. The University of Pretoria.

Mushi, G. E., Pienaar, H. and Deventer, M. van (2020) 'Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania', *Data Science Journal*, 19(1), pp. 1–9. DOI: https://doi.org/10.5334/dsj-2020-001.

Spinger Nature (2020) *Recommended repositories*. Available at: https://www.nature.com/sdata/policies/repositories#general (Accessed: 29 April 2020).

Tang, R. and Hu, Z. (2019) 'Providing Research Data Management (RDM) Services in Libraries: Preparedness, Roles, Challenges, and Training for RDM Practice', *Data and Information Management*, 3(2), pp. 84–101. DOI: https://doi.org/10.2478/dim-2019-0009.

Tenopir, C., Birch, B. and Allard, S. (2012) 'Academic Libraries and Data Services in Academic Libraries, *College & Research Libraries*, 46(3), pp. 61–75. Available at: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.

Yoon, A. and Schultz, T. (2017) 'Research Data Management Services in Academic Libraries in the US : A Content Analysis of Libraries ' Websites', *College & Research Libraries*, 78(7), pp. 1–19. Available at: https://crl.acrl.org/index.php/crl/rt/printerFriendly/16788/18346.

Endnotes

[1] Gilbert Exaud Mushi is an Assistant Lecturer and a Librarian at the Sokoine University of Agriculture in Tanzania. He works in the Directorate of Sokoine National Agricultural Library and can be reached at gilbert.mushi@sua.ac.tz.