RESEARCH PAPER

Identifying and Implementing Relevant Research Data Management Services for the Library at the University of Dodoma, Tanzania

Gilbert Exaud Mushi¹, Heila Pienaar² and Martie van Deventer²

¹ University of Dodoma, TZ
² University of Pretoria, ZA

Corresponding author: Gilbert Exaud Mushi (ray19mushi@gmail.com)

Research Data Management (RDM) services are increasingly becoming a subject of interest for academic and research libraries globally – this is also the case in developing countries. The interest is motivated by a need to support research activities through data sharing and collaboration both locally and internationally. Many institutions, especially in the developed countries, have implemented RDM services to accelerate research and innovation through e-Research but extensive RDM is not so common in developing countries. In reality many African universities and research institutions are yet to implement the most basic of data management services. We believe that the absence of political will and national government mandates on data management often hold back the development and implementation of RDM services. Similarly, research funding agencies are not yet applying sufficient pressure to ensure that Africa complies with the requirement to deposit research data in trusted repositories. While the context was acknowledged the University of Dodoma library staff realized that it is urgent to prepare for the inevitable – the time when RDM will be a requirement for research funding support.

This paper presents the results of research conducted at the University of Dodoma, Tanzania. The purpose of the research was to identify and report on relevant RDM services that need to be implemented so that researchers and university management could collaborate and make our research data accessible to the international community.

This paper presents findings on important issues for consideration when planning to develop and implement RDM services at a developing country academic institution. The paper also mentions the requirements for the sustainability of these initiatives.

Keywords: Research data management; research data; data curation; university library; University of Dodoma; Tanzania

1. Introduction

The changing nature of research, which is moving from the physical or traditional to a digital research environment, has made it essential to plan for research data storage, data use and re-use as well as for collaboration with peers. The use of advanced technology helps to easily collect, organize and analyse even large datasets for research. In could therefore be assumed that research data management (RDM) would be seen, by all, as a core function in this emerging research environment but to date, RDM has mostly been required by research funding agencies. Generally, the requirement is that grant beneficiaries should preserve, share and make data accessible for re-use. For funders the emphasis on RDM is, to ensure the efficient use of the available funds and then the acceleration of research and innovation through the rich collaboration of researchers.

The National Science Foundation in the United States of America (USA), the Australian National Data Service in Australia [now the Australian Research Data Commons] and the e-Science Core
Programme in the United Kingdom (UK) have all been involved in enforcing mandates and advocating for national legislative instruments on data retention and frameworks on responsible conduct of research (Chiware & Mathe 2016).

In Africa, the National Research Foundation in South Africa has enforced the retention of data for its funded research and provided a framework for data management services by academic and research institutions there. However, universities and research institutions in many other African countries (and most probably also other developing countries) are yet to implement data management services. The absence of research funding agency policies and national government mandates on data management further obstruct the development and implementation of RDM services.

In Tanzania, only research data transfer procedures were developed in 2010 to enable data transfer between Tanzania and foreign institutions but there is no policy on data sharing.

The recent study conducted at the University of Dodoma, Tanzania on identifying relevant RDM services, surfaced the need for researchers and university management to collaborate and make their data accessible to the international community. This paper presents major findings on important issues for consideration when planning to develop and implement RDM services at an African university. The paper also includes information on how to sustain the initiative once it has been initiated.

2. Defining Research Data Management

The main goal of RDM services is to ensure that maximum value is gained from the data collected in an investigation or research project. This is done by providing access to the data for use and re-use in the long-term (Chiware & Mathe 2016).

RDM, for this paper, is a phrase that includes all activities associated with a research data lifecycle, from data collection to data dissemination and data preservation. The purpose of these activities is to give continued access to the data. Tenopir et al. (2012) also mentioned aspects such as data management planning, digital curation, metadata creation and data conversion, as essential components of RDM services, and therefore these are also included when we refer to RDM.

Lastly, RDM is a team effort. Research stakeholders, namely researchers, funding agencies, libraries and research management, all have roles to play towards effective RDM services. Davidson et al. (2014 in Chiware & Mathe 2016) also acknowledged ethics advisors and IT professionals as actors involved in RDM support services. The level of contribution from each of the stakeholders may vary from one institution to another, based on the approach used to support research.

3. Literature Review

The literature reviewed in preparation for this paper included case studies on how to implement RDM services at academic institutions. Important advice from these studies are presented and discussed below.

3.1. Acknowledging institutional context

Conducting a requirements analysis is an important aspect in establishing RDM services (Coates 2014). The literature consulted highlighted the need to develop RDM in the context of institutional culture and the associated unique environment. Coates, (2014) and Reed (2015), advised that organizations ought to conduct a requirements analysis at the point when the organization is thinking of establishing RDM at the institution. They also indicated that the process helps to determine how good RDM fits into the existing research practices, what services should be offered, and what resources related to research data are required. Mboera, (2015) pointed out the barriers to access and sharing of scientific data collected by researchers using public or donor funding. These barriers include scientific and technical; legal and policy; and institutional and management barriers. It is apparent that, gathering the views of an institution's research stakeholders (researchers, librarians, IT teams, and administrators) could facilitate the identification of their immediate needs and that an analysis of these needs would lay the foundation of effective RDM services.

Conducting a complete needs assessment is said to be both time consuming and resource intensive. It is therefore recommended that institutions review the needs identified by other institutions and then included these in an instrument that aims to also surface the unique RDM requirements for the institution (Erway et al. 2016). This process ensures that common components are developed by all institutions yet, it allows for the development of additional specific needs of the institution (Jones et al. 2013 and Erway et al. 2016). Thus, learning from international peers provides input for localising RDM in the context of institutional research practices (Van Deventer & Pienaar 2015).
3.2. Components and services to support RDM

Implementing both generic and institution-specific services allow for slight variances in RDM services from one institution to another while at least the most important generic services are implemented across all institutions. This paper presents these generic RDM services and components (identified from the literature) that should be adopted by all institutions, also those in developing countries. These aspects were kept in mind when the research was conducted at the University of Dodoma.

3.2.1. Institutional RDM policy and strategy

Jones et al. (2013) mentioned the importance of developing an institutional strategy and policy towards establishing a sustainable and achievable RDM service. The strategy has to show an understanding of the current position and provide a definition of the desired future position. The strategy should stipulate the objectives, surface the set of planned activities, and provide a roadmap for the implementation of the initiatives over a set period (Jones et al. 2013; Erway 2013; Coates 2014). Developing the strategy and set of targets largely depends on the environmental scanning (mostly a literature review) and an analysis of the expressed needs.

Extensive consultations are essential when developing institutional RDM policy. It is important to acknowledge the roles that different stakeholders play and to meet as many as is possible of the needs and requirements of all the stakeholders. This is an important step towards ensuring wide-spread buy-in and adoption of the policy. Finally, a clearly stated policy facilitates ease of use by all the various stakeholders. The policy should address data management planning, the management of active data, the selection of data for long term preservation and the accessibility of data through the use of catalogues and repositories. Each of these are described in more detail below.

3.2.2. Data management planning

Data management planning constitutes one of the important research funder requirements when applying for a research grant (Jones 2011, Jones et al. 2013; Fary & Owen 2013; Kelly & Hodgett 2015; Davidson 2016). Some research universities and organisations consider data management planning as an integral part of data management practices even without any supporting policies from government or a research funding agency (Liu & Ding 2016). Despite variations in funders’ specific requirements, many require data management plans (DMPs) when a research proposal is submitted. A DMP provides the researcher with an opportunity to reflect on the realistic requirements for a successful research project. The plan presents basic information on how the researcher will collect data and, how data will be shared and preserved, including any restrictions that may apply. The DMP, when shared, also allows other stakeholders to plan and ensure that the necessary infrastructure is in place when the research is initiated.

3.2.3. Managing active data

Facilitating data management services involves supporting data management during the active stage of research. It requires flexibility and functionality that allow researchers to store, access and share their data during research, especially when collaborating with others (Jones et al. 2013). Studying existing data storage practices in the institution (such as how much data is produced; where is the data stored and what backup facilities are in place, provide a stepping stone to address problems associated with managing active data. Many preliminary studies in institutions have surfaced challenges such as the elevated risk of data loss and security breaches in interim approaches (Universal Serial Bus (USB) drives, external hard drives, and local computer storage) to store research data. To address the data storage challenges, institutions appear to be investing heavily in infrastructure to provide data storage capacity for “free”. The process involves extending the storage capacity of existing facilities, by utilising High-Performance Computing (HPC) facilities or by using cloud-based storage (Jones et al. 2013; Van Tuyl & Michalek 2015).

3.2.4. Data preservation selection

Data that have long-term value should be preserved and made available for re-use in future research activities (UK Research Council cited in Jones et al. 2013). It is acknowledged that preserving data is a costly process; however, keeping everything is more expensive due to the increased volume of data production and access to internet services. Data ought to be selected based on the value they have to the institution and the general public. Criteria, for the selection of data with long term value, should be set by the institution. Especially data that align with the institution’s mission, non-replicable data, unique data, and data with historical value are important to the organisation. Deposit agreements should be put in place before data is ingested and the importance of these agreements should be emphasized through advocacy and proper guidance by the institution.
3.2.5. Data catalogues
Managing data for long-term preservation, facilitating access and re-use, would need the creation of a proper record of all the available research datasets (Jones et al. 2013). The datasets could be made openly accessible to allow discoverability and reuse of research data but this is not a pre-requisite for a catalogue. When data is not openly available, the organisation should consider making the metadata available. Future researchers could then identify that research has been completed and perhaps establish a collaboration agreement with the institution (Research Councils UK 2015). When data is made accessible, it is recommended that persistent identifiers, such as Digital Object Identifiers (DOIs), are used to ensure that the metadata as well as the research data would remain accessible for re-use (EPSRC 2011).

3.2.6. Data repositories
The repository component is in effect an extension of the data catalogues. Good RDM practice emphasises the use of data repositories to keep the selected valuable data for long-term use. The repositories can be built within the institution or the organisation could actively promote the use of external services to make data openly available (Jones et al. 2013). Making use of a combination of both internal and external approaches is also a possibility. Setting up a data repository for the institution would require the development of specific skills, assessing requirements and the development of a repository that will address the collective expectations of all different stakeholders.

3.2.7. Supporting RDM from the library
University libraries have supported research activities over an extended period, through many services they offer. The growing recognition of the importance of research data in many institutions around the world has brought the need for university libraries to manage digital data through incorporating RDM into research services offered (Flores et al. 2015).

University libraries are being actively involved from the early stages of establishing RDM by: conducting the needs assessment of the academic research community (Corrall et al. 2013), participating in the creation of an RDM policy that includes both funders and institutional requirements, advocacy, creating awareness, providing support through training and consultation services, developing data repositories and creating metadata for research data (Tenopir et al. 2012; Jones et al. 2013; Pinfield et al. 2014; SCONUL 2015; Chiware & Mathe 2016). Librarians are assisting researchers in creating DMPs, and guiding them to align their plans with institutional and funder data policy requirements (Corrall et al. 2013; Grynoch 2016; Antell et al. 2014; Krier & Strasser 2014; Whitmire et al. 2015). Lastly, librarians have a role to play in facilitating the discovery and dissemination of the metadata for datasets held by the institution (Krier & Strasser 2014; Searle 2011).

Finding out whether researchers and managers at the University of Dodoma shared the opinion, that the development of RDM services was essential, was the focus of research that was conducted (Mushi, 2017a). The outcome of the research led to the development of a strategy which is discussed here, in section 5.

4. Methodology
The core objective of this study was to, for the library, identify the RDM services that have to be designed and developed so that the librarians at Dodoma University could assist the university's researchers to make their data accessible to the national and international community (Mushi, 2017a).

Ethical clearance, for the research, was granted by the Universities of Pretoria and Dodoma. Detailed informed consent information was provided to all participants prior to conducting the research.

Qualitative research was conducted. A case study, as well as a survey method, were used for this study. The population of the study consisted of 14 respondents which comprised of six (6) postgraduate students, six (6) researchers and two (2) university managers (Director of Library Services and Director of Research and Publications) from the University of Dodoma (Mushi, 2017b). Purposive sampling technique was used to select the two university managers and snowball sampling technique for the postgraduate students and researchers. An online questionnaire, making use of Google forms, was designed to collect data from the postgraduate students and researchers while interviews were used for the university managers. Data collected were thematically analysed and the major themes were then aligned with the objectives of the study.

5. Discussion of Results and Recommendations
It was established that researchers and postgraduate students at the University of Dodoma were not applying RDM best practices. Neither were external forces, such as research funding agencies or the government, mandating the use of DMPs.
However, the findings also indicated that respondents recommended the establishment of RDM services at the university. Therefore, it was recommended that, as a first step, a workgroup, representing several research disciplines, should be established to develop a standard DMP template for use at the University of Dodoma. The purpose is to ensure a standard RDM practice at the university while the DMPs could also be used for planning purposes. Adopting a DMP developed by another organization would be economical, however; it is necessary to do some customization of the template to align with the University’s objectives and the researchers’ culture and context.

Obviously, once the DMP template has been established, the development of appropriate training materials will be required. The training is a first step towards ensuring that the practice of completing a DMP before conducting research becomes a standard practice. Standard practice is essential for reliable planning of appropriate infrastructure and especially the necessary support services.

Further recommendations relate to the role of the university library in supporting RDM and the sustainability of the RDM initiative at the University of Dodoma. Given the context it was possible to develop a phased approach to RDM implementation and to make it clear what aspects the library would take responsibility for.

5.1. Implementing RDM at the university

The anticipated benefits, the changing nature of scientific research, national government and funding agency mandates were all considered key factors for the university to plan and implement RDM services. It was accepted that RDM would need to be implemented in several phases and a recommended strategy was developed (see Figure 1) and presented to the university. It was advised that four phases, described below, were used for implementation. Where necessary, the phases could overlap because of different stakeholders driving the phase. The strategy does take the components discussed in section 3.2 into consideration. Implementation was deliberately planned like this to allow the library to continue with activities before a policy is formally accepted.

Phase 1: Strategy, policy, procedures and infrastructure

This phase of RDM establishment lays down the foundation and research data management infrastructure in the context of the university through developing an RDM policy, and strategic plans for sustainable RDM services. A comprehensive RDM policy should clearly identify the responsibilities of each stakeholder and ensure that the project is seen as part of normal university activities. RDM involves different departments which include library services, the ICT department, the research office, researchers themselves as well as executive management. Each of these stakeholders have a role to play. Each role must be clearly stated in

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Strategy, policy, procedures and infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Create the essential documentation and infrastructure for the full scale roll-out of RDM. Establish storage facilities and create repositories.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 2</th>
<th>Awareness creation, skills develop and repository content development</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concentrate on training and advice. Develop skills for both librarians and researchers. Provide guidance for repositories and data storage. Establish a data catalogue.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 3</th>
<th>Management of active data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Make central storage available for active data. Enable the sharing of active data for collaborative projects.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 4</th>
<th>Data selection and preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Develop selection criteria, develop deposit agreements and licencing terms and conditions. Develop a preservation strategy and policy. Ingest relevant datasets.</td>
</tr>
</tbody>
</table>

Figure 1: Planned implementation phases.
the policy to ensure active participation and ultimately also project success. For this to be successful all stakeholders will be invited to participate in planning the policy.

The phase also involves the actual implementation of RDM infrastructure such as data storage, data repositories and developing important tools such as a data management planning tool. For the library this phase requires guiding, training, and supporting different stakeholders in performing their responsibilities. The library is expecting to participate but not to lead activities linked to this phase.

Phase 2: Awareness creation, skills develop and repository content development

The study indicated that University of Dodoma researchers and stakeholders have a low level of awareness of RDM practices. Therefore, a concerted effort is required to create awareness regarding the benefits of good RDM practices. Both researchers and university research management stakeholders need to be made aware of various issues related to RDM (such as funder mandates, accessing and sharing data, infrastructural requirements and support available at the university library).

It is urgent that the library ensures skills development for both librarians and research staff. There are very many online training modules available to initiate the activity but at some stage the library would need to also customise the training offering to make provision for the context and adapted requirements of the University of Dodoma specifically.

Some of the methods of creating awareness and promoting RDM practices include publishing articles via the university and library websites, symposiums, introducing the matter at the important institutional meetings, and through training and workshops.

This phase requires leadership and project management from the library. The library could expect participation from other stakeholders but these activities could be regarded as the library’s responsibility.

Phase 3: Management of active data

The focus in this phase is to manage active data which involves developing services to facilitate data management during the active stage of research. When the survey was conducted, it was established that 75% (six of the eight respondents) used personal computers and laptops; USB drives, external hard drives, CDs and DVDs to store their research data. Two of the participants were using cloud storage such as Dropbox, Google drive and OneDrive to store their research data. None of the respondents indicated the use of managed data repositories or central storage of research data. This phase therefore, includes ensuring that sufficient data storage capacity that will enable researchers to store, access, and share their research data during collaboration. It also does require that researchers are made aware of storage facilities in the cloud that are more reliable.

The library is expecting to participate but not to lead activities linked to this phase. It is anticipated that change management activities would need to be planned and that this phase would have to be well underway before the last phase could be implemented.

Phase 4: Data selection and preservation

The phase addresses issues related to identifying those datasets that are irreplaceable, data with high value. The development of data selection criteria is core to this phase because it will not be possible to retain all data. Similarly deposit agreement(s), deposit tools, and the development of guidance documentation also need attention.

This phase requires leadership and project management from the library but it is anticipated that the phase will not be initiated before phases one and two are well understood and the required outputs are in place. On the other hand, the library could expect participation from other stakeholders but these activities could be regarded as the library’s responsibility. If the library still has, by then, not developed the necessary skills to do so, knowledgeable experts would have to be contracted in to ensure that all valuable datasets are curated for the longer term.

5.2. Sustainability of RDM services at the university

An RDM implementation project is only as good as the associated sustainability plan. To most institutions, also the University of Dodoma, RDM represents a completely new programme of work that will bring about both organisational and behavioural changes. Creating and accepting a long-term strategy for RDM would surface the resources needed so that sustainability could be ensured by the institution.

Sustainability is seen as part of the initial stage (phase 1) of the RDM implementation project While RDM was accepted as a necessity the actual costs and expenditures still have to be determined. Clearly stating
RDM long-term goals, while keeping in mind the institutions’ mission and considering these in the context of the cost implications will provide an opportunity for university management to show commitment towards securing the resources for the sustainability of the services. University management is currently considering the financial implications.

In the absence of a sustainability plan the library still has a responsibility to develop the skills of the library staff and to support researchers with relevant services. As part of the sustainability of the second phase, described above, it is essential that the library remains actively involved by identifying alternative, often free, services and resources that are available to individual researchers.

6. Conclusion
Not having an RDM policy in place at the institution is no excuse for any library to not develop RDM skills nor not to provide RDM services. By considering the RDM implementation phases described in section 5.1, RDM components and services, identified from the literature as good practice, could be developed and systematically implemented by the University of Dodoma library. In doing so it will raise awareness within the institution and in the longer term it will ensure that the university’s research data is made accessible to the international community.

Acknowledgements
Sincere gratitude goes to the University of Pretoria and Carnegie New York Cooperation for funding the studies that led to this paper.

Competing Interests
The authors have no competing interests to declare.

References
References


Mushi, GE. 2017a. Identifying relevant research data management services for the University of Dodoma, Tanzania. Master of Information Technology(B), University of Pretoria. Unpublished.

Mushi, GE. 2017b. 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.


