Researchers' Perceptions of Research Data Management Activities at an Academic Library in a Developing Country

Johnson Mulongo Masinde, Central China Normal University, Wuhan, China

Jing Chen, Central China Normal University, Wuhan, China

Daniel Wambiri Muthee, Kenyatta University, Kahawa Sukari, Kenya

ABSTRACT

This study investigated researchers' perceptions of RDM activities at the Mahatma Gandhi Graduate Research Library (MGGL) to recommend measures to enhance managing, sharing, and reusing research data. The study was underpinned by the DCC curation lifecycle model and the community capability model framework (CCMF), which enabled the investigator to employ a descriptive research design to capture data from a broad cross-section of 330 researchers purposely. The data was analyzed using SPSS to generate descriptive and inferential statistics that enabled the investigator to address the research problem. Though the library had policies on research data, quality assurance, and intellectual property, study findings evidenced no explicit policies to guide each stage of data curation and capabilities. There were also inadequacies in the knowledge and skills capability, technological infrastructure, and collaborative partnerships. Overall, RDM faced challenges in all examined capabilities. The study recommends establishing an RDM unit within the library to oversee the implementation of RDM activities.

KEYWORDS

Academic Research Libraries, Emerging Economies, Research Data, Research Data Management Activities, Researchers' Perceptions

INTRODUCTION

Research data is ordinarily complex, irreplaceable, costly, and laborious to generate. They are valuable knowledge assets in an electronic or non-digitized form that research organizations should manage since they are the primary research materials generated by researchers during the research process (Ray, 2014). Accordingly, it is indispensable for research Institutions to effectively capture, describe, manage, and make available the data for discovery, sharing, and reusing. The propensity to manage, preserve, discover, share and reuse research data has become vital in furthering knowledge and science, augmentation of novel solutions to social and economic impediments, and amplifying immense potentials for competitiveness and productivity, and liveability (Tenopir et al., 2020).

Researchers are turning to academic libraries to manage their data owing to paradigm shifts in scientific research that are being propagated by cyberinfrastructures, funder mandates for research data sharing, and the proactive role of libraries in shaping scholarly communication. In this regard, the present study attempts to examine researchers' perceptions of research data management activities

at an academic library in an emerging economy to recommend measures to enhance managing, sharing, and reusing research data. The framing of the research problem is motivated by the need to ascertain the contemporary state of RDM in academic libraries of emerging economies from a user perspective. At the core of this was identifying the presence of the necessary capabilities to support efficient managing, sharing and reusing research data. The research approach undertaken enabled the research problem to be addressed in such a way as to permit comparisons over time.

Research Problem

RDM activities in most academic libraries of emerging economies have been described as lacking, rudimental, and unstructured (Chigwada, Chiparausha, & Kasiroori, 2017). Similarly, very little research was conducted to examine how the academic libraries capture, describe, manage, and make available research data for discovery, sharing, and reusing (Nhendodzashe & Pasipamire, 2017). The literature review also shows that not much attention was paid to key stakeholders such as researchers, librarians, the IT office, the research office, and the legal office. Therefore, this study investigates researchers' perceptions of RDM activities at an academic library in an emerging economy to recommend interventions to enhance managing, sharing, and reusing research data.

LITERATURE

Librarians and researchers reckon that the preservation and management of research data should be a key role in libraries. Banking on the reputable grounds of institutional repositories, academic libraries are expected to amass multiple data from scientific research activities and processes, select, classify, index, preserve and avail it for discovery, sharing, and reuse. Cox and Pinfield (2014) in a report titled 'The Last Mile' advance that university libraries have a duty to employ novel stratagems and assemble capabilities in the form of policies, expertise, technologies, and partnerships to enable efficient managing, sharing and reuse of data.

When research data is appropriately managed, it can ensure more productive and efficient science as new knowledge is created by building on discoveries, innovations, and open scientific inquiry (Nhendodzashe & Pasipamire, 2017). Furthermore, research data reuse can reduce the costs of researching by making data readily available while limiting its duplication and acting as a way of meeting funder mandates (Woeber, 2017). Universities, particularly in developed countries, have been addressing this strategic objective. Key RDM stakeholders, for example, academic libraries, the legal office, the IT office, the research office, and the mother institution, have been engaged in developing capabilities such as supportive policies, knowledge and skills, technological infrastructure, and started collaborating with other relevant institutions and organizations to ensure efficient managing, sharing, and reusing research data (Cox Andrew, 2019). In contrast, Ng'eno (2018) aver that research institutions in emerging economies have mainly restricted their efforts in RDM to advocacy programs, with most institutions lacking RDM activities or at the infant phase. In this regard, the researchers have repeatedly had to cope with multifaceted realities of making decisions in the absence of harmonized data management systems; which has proved challenging. Similarly, Nhendodzashe and Pasipamire (2017), found that the libraries lack the needed institutional capabilities such as supportive laws and policies, human resource capabilities, robust technological infrastructure, and collaborative partnerships to support RDM activities. In Kenya, a study by Anduvare (2019) on data support services among Kenyan, private universities observed that researchers managed their data that was in different formats and scattered all over the shelves and computers. There was no framework to guide RDM activities. Information professionals were deficient in RDM competencies and were not engaged in any form of RDM activities despite knowing its significance (Anduvare, 2019). Accordingly, the research institutions played little to no role in managing, sharing, and reusing research data, which resulted in the loss of valuable knowledge assets, duplication of findings, lack of

innovations, increased research costs, inability to meet funder mandates, and poor decision making (Anduvare, 2019).

To check RDM challenges, Flores, Brodeur, Daniels, Nichools, and Turnator (2015) propose an organizational change within the academic libraries, to establish a specific unit to oversee the managing of research data activities. According to the study, the unit should guide the whole research data management process, including the roles, responsibilities, services, and coordination. Such functions and responsibilities should include the formulation of RDM strategies and mechanisms, implementation and effecting of policy guidelines, and operationalization of awareness and advocacy campaigns (Flores et al., 2015). Additionally, the unit should develop capture instruments, clean and validate research data, design database systems and facilitate researcher collaboration at national and international levels. Further still, Flores et al. (2015) mention the need for the unit to manage the portal services, metadata standards, and all the data curation activities, including developing an ethical framework to guide the management of research data.

THEORIES UNDERPINNING THE STUDY

The Open Archival Information System (OAIS), the Maturity Model, the DCC Curation Lifecycle Model, and the CCMF have attempted to explain RDM. The OAIS does not include a phase before the ingest and pre-ingest function making it short of the theoretical basis to effectively underpin the study (Lavoie & Coalition, 2014). Similarly, the maturity model may be taken to infer a specific development course, resulting in a fixed and established definitive setting. However, this is rarely the case. A term like "underdeveloped," or "immature" when linked to maturity frameworks, may be regarded as disapproving (Cox Andrew, 2019). In this study, the theories that align well with the research problem are the CCMF and the DCC Curation Lifecycle Model.

The CCMF provides details of the roles and responsibilities of each capability for the effective management of research data. However, the model does not focus in detail on data curation which is the nucleus of RDM (Lyon, Ball, Duke, & Day, 2012). Similarly, the DCC Curation lifecycle model offers high-level view of data curation allowing the activities to be planned at different granularity levels including defining tasks and responsibilities; building frameworks of standards and technologies; and ensuring processes and policies are adequately documented (Higgins, 2012). However, like the CCMF, the model does not support RDM capabilities. Higgins (2012) advances that the model can be utilized in conjunction with other relevant models to plan and elucidate RDM. This investigation adopts the DCC Curation lifecycle model and the CCMF to explain researchers' perceptions towards RDM activities in an emerging economy. Methodology.

The investigation adopted a descriptive research design that was supported by the DCC Curation Lifecycle Model and the CCMF. The DCC Curation Lifecycle Model underpinned data curation activities while the CCMF informed capabilities (Policy guidelines, knowledge, and skills, technology and infrastructure, and collaborative partnerships). According to Sekaran (2003), a descriptive research design is conducted to describe constructs of interest in an investigation. It is suitable when a researcher wants information on persons' insights, expectations, experiences, perceptions, etc. The design was employed to ascertain and describe perceptions of researchers towards RDM activities at the MGGL.

LOCATION OF THE STUDY

The MGGL is part of University of Nairobi's Library System and stands as a hub for providing research services to postgraduate students and academic staff. The research library offers a suite of research data services (RDS), including partnerships in the form of linkages and workgroups, interdisciplinary and emerging research methods, and expert guidance in research processes. The MGGL is also manned by expert personnel with competencies in research-related services. The study selected the MGGL because: i. In 2020, the US news Education ranked the University of Nairobi position 792

globally, 18th in Africa, and 1st in Kenya depicting its research standing and suitability as a case (U.S.News & World Report, 2020). ii. The University of Nairobi has an annual research kitty of 45 million USD, the highest in East Africa. The funding signifies its involvement in rich, valuable, and high intensive research (Nacosti, 2020). iii. It has the largest student population in Kenya, with more than 84,000 students and at least 14,500 postgraduate students (University of Nairobi, 2020). iv. It was the first to be ISO certified in 2008 (University of Nairobi, 2015). With ISO certification comes amplified expectations in quality, customer care, and satisfaction. Therefore, the researchers expect efficient and high-quality RDS.

Target Population

The investigation targeted the researchers (postgraduate students and academic staff) who were conversant with RDM activities at the MGGL.

Sampling Techniques and Sample Size

The study considered it appropriate to select participants through purposive sampling. The University of Nairobi has 14567 Postgraduates 2,220 academic staff with PhDs and 450 professors totaling a target population of 17237 researchers. The sample size was calculated by determining the sample size of the infinite population and then adjusting it to the sample size of the required population (Cochran, 1977).

Formulae of sample size for infinite population:

$S = Z^2 * P * (1-P) / M^2$
Confidence level = 95%
Z = 1.96
M = 0.05
S = 384.16

Adjustment of the sample to the required population:

Formulae for Adjusted Sample size = (S)/ 1+ [(S – 1)/ Population]
Sample size= 384

DATA COLLECTION

The investigation utilized a questionnaire. The questions were designed to explore subjects reflecting the research problem and were developed using information gained from the literature review, contacts in the field, and for comparative purposes with prior surveys.

PILOT STUDIES

The investigation piloted the instruments among researchers at the KU Post Modern Library. The Investigator chose a different library to check the contamination of the study. The KU Post Modern Library was selected because it has similar characteristics to those of the MGGL.

The number of participants in the pilot was 17 (12 postgraduates and 5 Academic staff). The number was informed by Hill and Hamilton (1998), who suggested samples of between 10-30 for pilot studies. After the pilot, the Investigator invited experts for further assessment. The outcome was then used to test the instruments' internal consistency using the Cronbach alpha, which produced an

Table 1. Cronbach's alpha coefficients

Variable	Items	Cronbach's alpha
Policy guidelines	13	0.876
Digital curation activities	6	0.728
Knowledge and skills	13	0.845
Technology and infrastructure	5	0.721
Collaborative partnerships	3	0.809

average above 0.7 for all constructs. Accordingly, the instrument was deemed reliable and acceptable. Table 1 shows Cronbach's alpha coefficients.

DATA COLLECTION PROCEDURES

The Investigator contacted the researchers between February and May 2020 for a briefing and subsequent invitation to participate in the study.

DATA ANALYSIS

Before data analysis, questionnaires were checked for correctness by examining whether they had been properly and completely filled. The Investigator utilized the Statistical Software for Social Sciences (SPSS) v 21.0 to analyze the data quantitatively.

ETHICAL CONSIDERATIONS

Ethics in research warrant that no individual suffers the harmful consequence of research activities (Masinde, Wambiri, & Chen, 2020). Consequently, the Investigator took time to inform respondents of the nature and aim of the research before requesting their consent to participate in the study.

DATA ANALYSIS AND DISCUSSION

Return Rate

Out of the 384 questionnaires sent to researchers, 330 were returned, representing an 86% return rate. According to Morton, Bandara, Robinson, and Carr (2012), a return rate approximating to 60% for most studies is considered good, while a 50% return rate is suitable for analysis. A 70-85% return rate is very good, while a return rate above 85% is regarded as excellent. Therefore, the response was regarded as satisfactory and acceptable.

Demographic Characteristics

Consequently, 330 questionnaires were analyzed. Of the 330 participants, 171(51.82%) were male, and 159(48.18%) were female. These outcomes indicate that almost an equal number of each sex participated in the survey. It was also revealed that 11(3.33%) of the participants were Professors, 89(26.97%) were Doctors, 146(44.24%) were Ph.D. students while 9(2.73%) were Master's degree students. 75(22.73%) of the respondents were taking other Postgraduate programs. Besides, 42(12.73%) were aged between 18-35 years, and 169(51.21%) were aged between 36-45 years. There were also 119(36.06%) participants aged above 45 years. The data is illustrated in Table 2.

Table 2. Characteristics of participants

Participants' Characteristics		Frequency	Percentage
Gender	Female	159	48.18
	Male	171	51.82
Educational Level	Professor	11	3.33
	Doctor	89	26.97
	PhD students	146	44.24
	Master Students	9	2.73
	Other Postgraduate	75	22.73
Age	18-35	42	12.73
	36-45	169	51.21
	Over 45	119	36.06
Frequency of research service use	Weekly	252	76.36
	Monthly	57	17.27
	A few times a semester	17	5.15
	A few times a year	4	1.21

Participant characteristics, N=330

The study also investigated how frequently participants utilized RDS. Respondents were asked to choose between week, month, a few times in a semester, and a few times in an academic calendar. The study shows that participants were heavy consumers of RDS. The investigation pooled weekly and monthly users and categorized them as regular users. This group made up 93.8% of the participants. The participants who utilized RDS a few times a semester were categorized as occasional users and represented 5.15% of the users. The users who only utilized the services a few times a year were categorized as rare users and represented 1.21%.

The findings are presented in the context of the key themes as shown below:

1. Policy guidelines that influence RDM activities.
2. Digital curation activities.
3. Knowledge and skills influencing RDM activities.
4. Technology and infrastructure readiness for RDM activities.
5. Collaborative partnerships influencing RDM activities.

Policy Guidelines Influencing RDM Activities

Policy guidelines respond to RDM activities such as preservation, storage, quality, security, sharing, compliance, and jurisdiction to improve managing sharing and reusing research data (Nhendodzashe & Pasipamire, 2017). They are supported by the DCC Curation lifecycle model which defines roles and policies at each stage and the CCMF which ensures that processes and policies are adequately documented and adhered to at all stages after being defined and premeditated (Higgins, 2012; Lyon et al., 2012).

The Cronbach's alpha coefficient for this question was 0.876 signifying an acceptable internal validity (Gliem & Gliem, 2003). Consequently, the study investigated the researchers' perceptions of policy guidelines influencing RDM activities.

Table 3. Explicit policies

Explicit policies	Mean	Std. Deviation	N
Data capture	3.68	.881	330
Data appraisal	3.56	.770	330
Data description	3.64	.788	330
Data preservation	3.70	.875	330
Data access	3.59	.739	330
Data use and reuse	3.63	.830	330
Data sharing	3.69	.833	330

Policy Guidelines That Governed RDM Activities

Respondents were asked to reveal if the library had RDM policies such as the research data policy, the intellectual property rights policy, the quality assurance policy, a policy on Knowledge, skills, and training, a policy on technological infrastructure, and a policy on collaborative partnerships.

Study findings indicate that the MGGL had the Research data policy (Mean =2.37), the Intellectual Property rights policy (Mean =2.36), and the Quality assurance policy (Mean =2.28). However, the library did not have a policy on Knowledge, skills, and training (Mean =3.61), technological infrastructure (Mean = 3.65), and collaborative partnerships (Mean =3.72).

Explicit Policies on Data Curation Activities

The study further asked respondents to reveal whether the library had explicit policies on data curation activities such as data capture, data appraisal, data description, data preservation, data access, data use and reuse, and data sharing. The findings show that the library did not have explicit policies on data curation activities as shown in Table 3.

Investigation of Data Curation Activities

Data curation is the activity of active and continuing management of research data through its lifecycle of interest and usefulness for discovery, sharing, and reuse (Higgins, 2012). It is supported by the DCC Curation lifecycle model; which supports digital materials' management, permitting their effective curation and preservation from initial conceptualization to disposal or selection for use, reuse, and long-term preservation. The activities involved in data curation include data capture, appraisal, description, preservation, access, reuse, and transformation of research data (Higgins, 2012). Similarly, the CCMF imparts data curation activities by supporting capabilities (Policy guidelines, knowledge and skills, technology and infrastructure, and partnerships) in data curation (Lyon et al., 2012).

The Cronbach's alpha coefficient for this question was 0.872 indicating a high internal validity. The findings are presented below.

Capturing Data

The study asked respondents to state the tools they used to capture their data. Study findings show that the majority captured their data through questionnaires 115(34.8%), Interviews 104(31.5%), and audio recordings 91(27.6%) while a significantly smaller number captured their data through cameras 9(2.7%), geographic information systems (GIS) 4(1.2%), field experiments 4(1.2%), and laboratory experiments 3(0.9%). Study findings are shown in Figure 1.

Figure 1. Data capturing tools
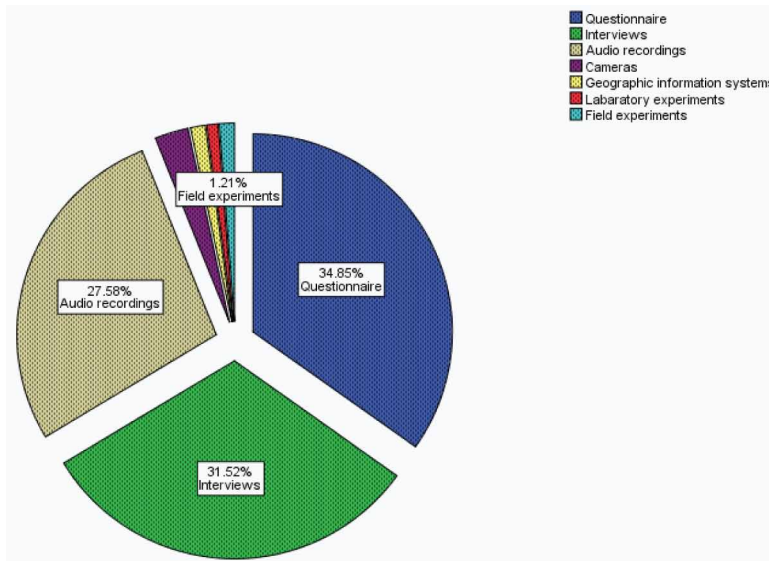
Volume of Research Data Generated

Respondents were similarly requested to reveal the volume of data they generated per project. Study findings demonstrate that the majority 303(91.8%) generated between (1-50GB) while a considerably smaller number 15(4.5%), 9(2.7%) and 3(0.9%) generated between (50-100GB), (101GB-1TB) and (1TB-1PB) respectively.

Metadata Creation

Metadata is information that describes the characteristics of datasets making them discoverable, citable, and reusable (Higgins, 2012). The study investigated whether respondents generated their metadata or the library generated metadata for use.

The findings show that most respondents 284(86.1%) stated that the library generated metadata for use. However, a smaller number 44(13.3%) indicated that they generated their metadata.

Storage and Preservation of Research Data

A storage and preservation mechanism ought to be reliable, guarantee safety, permit discovery, and access to data and observe all the RDM best practices (Chigwada et al., 2017). Accordingly, the study investigated the short term and long term data storage.

Storage of Data on Short Term

Respondents were requested to state where they stored their data in the short term. Based on the findings, a majority142(43%) stored their data in their laptops while 94(28.5%), 46(13.9%), 40(12.1%), and 8(2.4%) stored their data in the external hard drive, cloud service, library repository, and other storage sources respectively.

Storage of Data on Long Term

Respondents were also asked to reveal where they stored their data on long-term. The findings evidence that the majority145(43.9%) deposited their data in the library repository. However, 83(25.2%),

74(22.4%), 25(7.6%) and 3(0.9%) stored their data in the external hard drive, cloud service, laptops, and other forms respectively.

Sharing and Reusing Research Data

Sharing and reusing research data ensures the development of new knowledge through building on previous research, innovations, and open scientific inquiry (Tenopir et al., 2020). The study similarly investigated the access, sharing, and reuse activities of researchers at the MGGL.

Sharing of Research Data

The researchers were asked to reveal if they were freely willing to share their data. Most respondents 238(72.1%) were unwilling while 77(23.3%) did not have any problem sharing their data. A substantially smaller number 15(4.5%) were indifferent.

Research Data Use and Reuse

Equally, the study investigated the reasons for the use and re-use of data among respondents. Study findings show that respondents used and re-used research data to validate the findings and reduce costs involved in research (Mean=1.46), Promote new discoveries and innovation (M=1.75), Create an enabling environment for new collaborations (M=1.80), Meet donor requirements (M=1.92) Increase the research impact (M=1.98), and Promote open scientific discovery, access, sharing and reuse (M=2.15).

Knowledge and Skills for RDM Activities

The DCC Curation lifecycle model, which underpins the study, sets out the roles at every stage of data curation. The model blends with the CCMF to support the knowledge and skills requirements for RDM activities.

The Cronbach's Alpha for the question was 0.845 signifying a high internal validity. Consequently, the study investigated researchers' perceptions of the RDM knowledge and skills.

Level of Knowledge You Possess Regarding RDM

Respondents were asked to rate the level of knowledge they possessed regarding RDM activities. Study findings show that a majority of 195(59.1%) rated their knowledge as average while 124(37.6%) rated their knowledge poor. However, a few of the respondents 11(3.3%) thought they had good knowledge of RDM.

Knowledge for Specific Data Curation Activities

Researchers were also asked about the knowledge they possessed regarding specific data curation activities such as data capture, appraisal, description, preservation, storage, access, sharing, reuse, etc. Research findings suggest that most respondents 172(52.1%) had poor knowledge of specific data curation activities while a smaller number 156(47.3%) rated their knowledge as average. Only 2(0.6%) of the respondents rated their knowledge of specific data curation activities as good.

Types of Skills You Possess Regarding RDM

The study further sought to determine the types of skills researchers possessed regarding RDM activities. The respondents were asked to state if they possessed skills in RDM activities such as metadata creation, research data discovery and retrieval, research data capture, research data preservation and sharing, RDM information technologies, and RDM collaborative networks.

The findings demonstrate that the majority 186(56.3%) possessed lesser skills in RDM activities such as metadata creation, research data discovery and retrieval, research data capture, research data preservation and sharing, RDM information technologies, and RDM collaborative networks while

a somewhat smaller number 143(43.3%) stated that they had average skills. Only one respondent 1(0.3%) indicated that they possessed skills in the RDM activities.

RDM Training Needs

The study investigated the urgent RDM training needs of researchers. Respondents were requested to reveal their immediate RDM training as regards to RDM activities such as training on metadata creation, data curation activities, RDM Information technology tools, and equipment, legal and ethical considerations in RDM, data management plan (DMP), grant proposal writing, collaborative partnerships and security and storage of research data.

Study findings evidence that respondents regarded all the RDM training needs as immediate. The respondents rated data curation training (85.8%), metadata training (84.9%), training on Information technology tools and equipment (81.8%), training on grant proposal writing (81.2%), training on security and storage of research data (80%), training on the DMP (78.2%), training on legal policies and ethical considerations (75.7%) and training on collaborative partnerships (74.8%) as immediate.

Quality of RDS

Similarly, the study investigated the quality of RDS offered by the MGGL. Respondents were requested to rate the quality of information data services and technical data services. The findings show that most respondents 186(56.3%) rated information data services (Maintaining a web resource guide, reference and advisory services, data publication advisory, RDM advocacy, and training, provision of support for search and retrieval of external data sources, directly participate with researchers on a research project, advisory services on data analysis/mining/visualization, etc.) dissatisfactory while a smaller number 118(35.8%) were indifferent. Few 26(7.9%) rated them satisfactorily.

On the technical data services offered, (run a data repository, advisory on the curation of active data, carry out long-term preservation of research data, accessioning and or deaccessioning datasets for deposit in the repository, creating metadata, provide a data catalog, offer advisory on copyright/licensing rights and intellectual property rights) a majority 204(61.8%) also rated the services dissatisfactory while 115(34.8%) were indifferent. A considerably smaller number 11(3.3%) thought the services were satisfactory.

Data Management Plan

Researchers were also asked to state whether they had a data management plan (DMP). The findings show that most respondents 242(73.3%) did not have a DMP while a fraction of 88(26.7%) revealed that they had a DMP.

Technological Infrastructure Readiness

Technological Infrastructure permits ease in accessibility and management of research outputs, creating more appropriate ways of propagating research data, consequently promoting knowledge integration (Schultz, 2017). The variable is supported by the DCC Curation lifecycle model which builds standards and frameworks, creating room for technological changes and ensuring the transition from each stage; and the CCMF which ensures all technical capabilities at each stage are executed adeptly.

The Cronbach's Alpha coefficient values for this question were 0.721. Accordingly, the study investigated researchers' perceptions of RDM technological infrastructure readiness at the MGGL.

Satisfaction With RDM Computing Facilities

Respondents were requested to rate their level of satisfaction with RDM computing facilities. A majority of 207(62.7%) rated computing facilities as dissatisfactory while 101(30.6%) were indifferent. A much smaller number 22(6.7%) rated them satisfactorily.

Figure 2. Satisfaction with software solutions
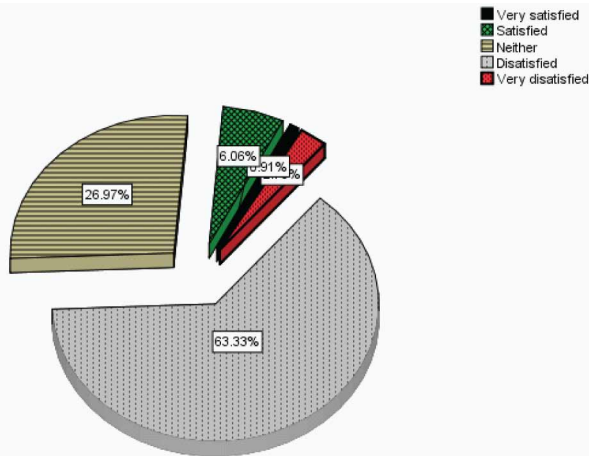
Satisfaction With RDM Software Solutions

Software such as MATLAB, Stata, R, SPSS, NVivo, etc. is used by researchers in research-related activities. The study investigated respondents' level of satisfaction with RDM software solutions. Study findings show that most respondents 218(66%) rated RDM software solutions as dissatisfactory while a smaller number 89(27%) were neither satisfied nor dissatisfied. However, few of the respondents 23(7%) rated the software solutions satisfactorily. The findings are shown in Figure 2.

Satisfaction With RDM Storage Media

It is critical to have the right and enough storage media for data storage and preservation. Consequently, the study investigated respondents' level of satisfaction with RDM storage media. The findings evidence that most respondents 177(53.6%) thought RDM storage media was dissatisfactory while 144(43.6%) were unsure. However, 9(2.7%) rated them satisfactorily.

Satisfaction With RDM Data Backup Tools

A backup is essential in RDM because it ensures that research data is restored after a loss. Therefore, the study investigated respondents' level of satisfaction with RDM data backup tools.

The findings demonstrate that the majority 169(51.2%) rated RDM data backup tools as dissatisfactory while 150(45.5%) were irresolute. 11(3.3%) rated the services satisfactory.

Satisfaction With RDM Security Solutions

Research data is generally sensitive and should be protected from unauthorized access and use. Therefore, the study investigated the respondents' level of satisfaction with RDM security solutions. The findings suggest that most respondents160(48.5%) were dissatisfied with RDM security solutions while a somewhat smaller number 143(43.3%) were not sure. On the contrary, 27(8.2%) of the respondents thought RDM security solutions were satisfactory.

Collaborative Partnerships Affecting RDM Activities

Collaborative partnerships between the library and departments or schools within the University, at the national and international level, allow and promote the implementation of shared policies, technical frameworks, documentation, workflows, and cost cuts augmenting access, sharing, and reusing data (Finkel et al., 2020). The DCC Curation lifecycle model underpinned the variable by shaping access,

Figure 3. Satisfaction with collaborative partnerships between the MGGL and other universities, Institutions, or organizations Internationally
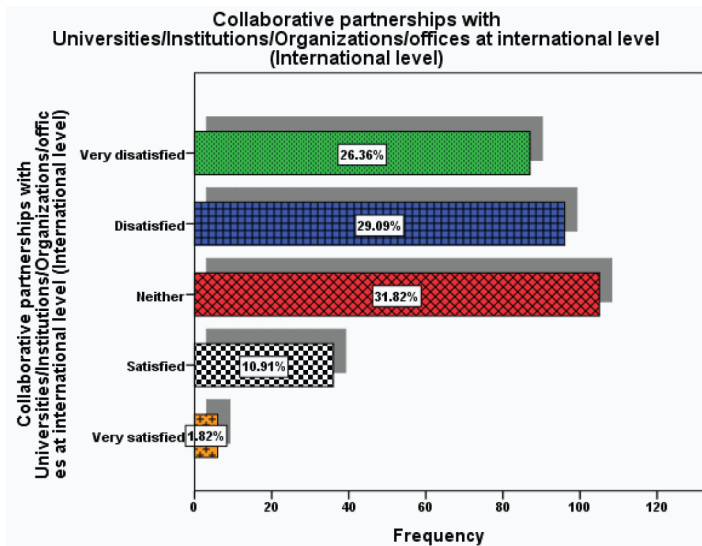
sharing, and reusing data. The CCMF focused on capabilities in collaborations between the library and departments or schools within the University, nationally and internationally.

The Cronbach's Alpha coefficient value for this question was 0.809. Consequently, the study investigated researchers' perceptions of the MGGL's RDM collaborations.

Satisfaction With Internal Collaborative Partnerships

Respondents were asked to reveal how satisfied they were with RDM collaborative partnerships between the library and other departments or disciplines within the University. Study findings indicate that a majority 142(43%) rated the collaborations dissatisfactory while 100(30.3%) were irresolute. However, a fraction of 88(26.7%) agreed that they were satisfactory.

Satisfaction With External Collaborative Partnerships

The study further investigated how satisfied respondents were with RDM collaborative partnerships between the library and other Universities, or organizations nationally. Study findings evidence that most respondents 160(48.5%) were dissatisfied with the collaborations while a smaller number 113(34.2%) were neither satisfied nor dissatisfied. However, a few of the respondents 57(17.3%) agreed that the collaborations were satisfactory.

Satisfaction With Collaborative Partnerships Internationally

Lastly, the study investigated how satisfied respondents were with RDM collaborative partnerships between the library and universities or organizations internationally. The findings indicate that a majority of 183(55.5%) rated the collaborations dissatisfactory while 105(31.8%) were irresolute. On the contrary, 45(12.7%) thought the services were satisfactory. The findings are shown in Figure 3.

POLICY GUIDELINES INFLUENCING RDM ACTIVITIES

Despite the presence of policies on research data, quality assurance, and intellectual property, study findings found no explicit policies to guide each phase of data curation and capabilities. These findings

contradict those by Nhendodzashe and Pasipamire (2017), who acknowledge that policies are the cornerstone of RDM. They guide RDM operationalization. The policies also build stakeholders' confidence by ensuring research data retention, storage, and availability for use, reuse, sharing, or disposal in line with the ethical, legal, statutory, and donor requirements (Chigwada et al., 2017). Section 2 of Australia's Code for the Responsible Conduct of Research states that policies are necessary to guide ownership, storage, and preservation of research data, retention after project completion, and access convenience (Australian National Data Service, 2016). Therefore, the management in consultation with key stakeholders must develop and fully implement explicit RDM policies. These policies will guide each phase of data curation and capabilities such as knowledge and skills, technological infrastructure, and collaborative partnerships.

DATA CURATION ACTIVITIES

At the heart of RDM are data curation activities, its basis. First, the study findings showed that researchers used different data capture methods with a questionnaire preference, an old data capturing technology. Firat (2017) observes that researchers must adopt real-time data capture methods to boost data integrity, leverage data automation, and minimize errors and costs. The findings further reveal that researchers generated data volumes of between 1GB and 50GB. The small data amounts indicate small-scale research activities synonymous with inadequate funding, as well as researchers having insufficient support to conduct highly intensive research (Chigwada et al., 2017).

Second, the study findings established that the library generated metadata for probable purposes of privacy, consistency, simplicity, and clarity in relationships (Higgins, 2012). However, one major challenge of uniform metadata is the inability to display all the metadata, which may affect discovery, sharing, and reuse. Ricardo, João, João, and Cristina (2015) explain that metadata must be created at the point of data capture to guarantee richness.

Third, study findings suggested that researchers preferred laptops in the short term and favoured repositories in the long-term. Kurata, Matsubayashi, and Mine (2017) enlighten that typically, researchers stay away from repositories in the short-term to check unethical data practices; they only considered the library after publishing their work. Chigwada et al. (2017) observe that it is critical to maintain standardized storage and frequent RDM training to equip researchers with knowledge on the significance of storage and preservation mechanisms.

Fourth, the study demonstrated that researchers were reluctant to avail data for reuse and sharing. Warren (2016) explains that key challenges on data sharing include deficiencies in supportive policies and technologies, researchers' hesitancy to share data, and inadequate data training and advocacy programs. Consequently, for useful and practical sharing and reuse, it is necessary to address issues on policy, privacy, and practicality.

Lastly, the findings revealed that the least consideration for using and reusing data was to promote open scientific discovery, access, sharing, and reusing. This finding suggests that researchers did not give extreme importance to open scientific discovery, access, sharing, and reusing because of data sharing reluctance. Overall, researchers were unwilling to share data because of data abuse through unethical practices, deficiencies in supportive policies and fear of losing competitive advantage (Kurata et al., 2017). Tenopir et al. (2020) encourage research data advocacy and training to advance research data access, sharing, and reuse.

KNOWLEDGE AND SKILLS FOR RDM ACTIVITIES

The study established that researchers had deficiencies in RDM knowledge and skills. knowledge and skills are considered a key component of RDM because they promote its direct and indirect benefits (Cox Andrew, 2019). The MGGL can conduct RDM training, advocacy, and support programs in data curation, metadata construction, information technology tools and equipment, grant proposal

writing, security and storage of data, DMP, legal policies and ethical considerations and collaborative partnerships. RDM training may be offered through workshops, online courses, lectures, and tutorials. The University of Manchester Library provides a comprehensive RDM training and advocacy program dubbed "My Research Essentials" (MRE). It uses workshops, information sessions, and online resources designed to promote RDM and support researcher development at different career stages. The program covers topics including "why researchers should share and reuse research data," "research data discovery," "the research process," "raising individual research profile," et cetera (The University of Manchester Library, 2020).

Quality of RDS

Study findings found that most researchers rated both information and technical data services as dissatisfactory. This finding is corroborated by Dér (2015), who indicates that RDS in academic libraries of emerging economies are basic and unstructured. Dér (2015) attributes the poor and rudimental services to a lack of supportive policies and technological infrastructure, human resources, and insufficient funding. Flores et al. (2015) recommend establishing a division within the library to guide RDS implementation by assembling all the needed RDM capabilities.

Data Management Plan

The majority of researchers did not have DMP. A DMP addresses key RDM areas such as how the researcher plans to collect, appraise, describe, preserve and avail data for sharing and reuse. Cox Andrew (2019) points out that academic libraries needed to regularly conduct advocacy and data training programs to furnish researchers with the knowledge and skills in managing, sharing, and preserving research data. Accordingly, the library should proactively promote and equip the researchers with knowledge and skills in writing a DMP.

TECHNOLOGICAL INFRASTRUCTURE READINESS

Study findings evidenced deficiencies in the items of technological infrastructure capability. Deficiencies were apparent in computing facilities, software solutions, storage media, data backups, and security solutions. For this reason, vast and robust technologies are necessary to support the whole research data lifecycle effectively. Similarly, the computing facilities and tools should have the capacity to support the storage and backup, preservation, and security of research data, including other RDM computing capabilities, such as interoperability, reliability, accessibility, and request responsiveness. According to the Helsinki University Library (2020), the RDM computing technologies must be robust and effective in data management and planning, description and documentation, security and ethical subjects, storage and backups, accessibility, and research data preservation.

COLLABORATIVE PARTNERSHIPS AFFECTING RDM ACTIVITIES

The study established that researchers were generally dissatisfied with collaborative partnerships. In view of that, the library should move to foster collaborative partnerships internally, externally, and internationally. Internally, the library should engage stakeholders to develop policies to guide collaborations. The library should explore other areas of collaboration including shared policies and documentation, workflows, and cost cuts with departments such as the information technology department, the research office, the legal office, and the academic departments. At the national level, the library, in consultation with the relevant authorities, should champion the establishment of an RDM collaborative framework through legislation and policy to coordinate and guide joint activities. The framework should drive the creation of opportunities for engagement and mutual benefits,

such as funding, scholarships, knowledge exchange, and job creation internally, at the national and international levels.

CONCLUSION

Academic libraries are moving from custodians to creators of knowledge resources through scholarly communication. This has particularly been typified in academic libraries of developed countries. However, study findings demonstrate that academic libraries in emerging economies are either lacking or still at the infant stages. For instance, RDM activities at the MGGL faced challenges in all the examined capabilities.

Recommendation

Like its counterparts in developed countries, the study recommends an organizational change, to develop an RDM unit within the MGGL to guide RDM implementation by assembling all the needed capabilities (Policy guidelines, knowledge and skills, technological infrastructure, and collaborative partnerships) to enable efficient managing, sharing, and reusing research data.

Implication

This present study adds to existing literature from an emerging economy perspective. There is very little research on RDM in emerging economies, especially in the Kenyan setting. The study also has the potentials of influencing the implementation and efficient managing, sharing, and reuse of research data in Kenyan and other academic libraries of emerging economies.

Future Research

The study investigated researchers' perceptions of RDM activities. Future studies could involve other key RDM stakeholders such as research librarians, the legal office, IT personnel, and the research office to permit decisive and far-reaching conclusions. Future studies could also investigate the presence of RDM legal framework.

REFERENCES

Anduvare, E. M. (2019). *eResearch Support: An Exploratory Study of Private University Libraries in Nairobi County, Kenya.* University of KwaZulu-Natal. Retrieved from https://pdfs.semanticscholar.org/12db/fde09819dda35be1b99eba72d71d5fbdfbe1.pdf

Australian National Data Service. (2016). *Australian Code for Responsible Research*. Australian Government. Retrieved from https://www.nhmrc.gov.au/guidelines-publications/r39

Chigwada, J., Chiparausha, B., & Kasiroori, J. (2017). Research Data Management in Research Institutions in Zimbabwe. *Data Science Journal*, *16*(0), 16. doi:10.5334/dsj-2017-031

Cochran, W. G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley & Sons Ltd.

Cox, A., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299–316. doi:10.1177/0961000613492542

Cox Andrew, M. (2019). Maturing research data services and the transformation of academic libraries. *The Journal of Documentation*, *75*(6), 1432–1462. doi:10.1108/JD-12-2018-0211

Dér, Á. (2015). *Exploring the academic libraries' readiness for research data management: cases from Hungary and Estonia*. Oslo and Akershus University College of Applied Sciences.

Finkel, M., Baur, A., Weber, T. K. D., Osenbrück, K., Rügner, H., Leven, C., Schwientek, M., Schlögl, J., Hahn, U., Streck, T., Cirpka, O. A., Walter, T., & Grathwohl, P. (2020). Managing collaborative research data for integrated, interdisciplinary environmental research. *Earth Science Informatics*, *13*(3), 641–654. doi:10.1007/s12145-020-00441-0

Firat, H. K. M. (2017). Opinions of Expert Academicians on Online Data Collection and Voluntary Participation in Social Sciences Research. *Educational Sciences: Theory and Practice*, *17*(5), 1461–1486.

Flores, J. R., Brodeur, J. J., Daniels, M. G., Nichools, N., & Turnator, E. (2015). Libraries and the Research Data Management Landscape. *Council on library and Information resources*, 82 - 102.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, Interpreting, And Reporting Cronbach s Alpha Reliability Coefficient For Likert-Type Scales*. Academic Press.

Helsinki University Library. (2020). *Research data management*. Retrieved from https://www.helsinki.fi/en/research/services-for-researchers-and-research-policy/research-data-management

Higgins, S. (2012). The lifecycle of data management. In G. Pryor (Ed.), *Managing Research Data* (pp. 17–46). Facet. doi:10.29085/9781856048910.003

Hill, R., & Hamilton, W. P. (1998). *What sample size is "enough" in internet survey research?* Academic Press.

Kurata, K., Matsubayashi, M., & Mine, S. (2017). Identifying the Complex Position of Research Data and Data Sharing Among Researchers in Natural Science. *SAGE Open*, ●●●, 7.

Lavoie, B. F., & Coalition, D. P. (2014). *The Open Archival Information System (OAIS) Reference Model: Introductory Guide* (2nd ed.). Digital Preservation Coalition.

Lyon, L., Ball, A., Duke, M., & Day, M. (2012). Developing a Community Capability Model Framework for data-intensive research. *Proceedings of the 9th International Conference on the Preservation of Digital Objects*.

Masinde, J. M., Wambiri, D. M., & Chen, J. (2020). *Gender and cognitive factors influencing information seeking of graduate students at Kenyatta University Library*. 10.4102/sajim.v22i1.1154

Morton, S., Bandara, D. K., Robinson, E., & Carr, P. A. (2012). In the 21st Century, what is an acceptable response rate? *Australian and New Zealand Journal of Public Health*, 36. PMID:22487341

Nacosti. (2020). *National Research Priorities 2018-2022*. Nairobi, Kenya: National Commision for Science, Technology and Innovation.

Ng'eno, E. J. (2018). Research Data Management in Kenya's Agricultural Research Institutes. *African Journal of Library Archives and Information Science*.

Nhendodzashe, N., & Pasipamire, N. (2017). *Research data management services: are academic libraries in Zimbabwe ready? The case of University of Zimbabwe library.* Paper presented at the IFLA.

Ray, J. E. (2014). Research Data Management Practical Strategies for Information Professionals. Purdue University Press.

Ricardo, A., João, A., João, R., & Cristina, R. (2015). A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential. *Advances in Intelligent Systems and Computing*, *353*, 101–111. doi:10.1007/978-3-319-16486-1_10

Schultz, A. Y. T. (2017). Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. *College & Research Libraries*, *78*(7).

Sekaran, U. (2003). Research Methods for Business: A Skill-Building Approach (4th ed.). New York: John Wiley & Sons.

Tenopir, C., Rice, N., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS One*, *15*(3), 15. doi:10.1371/journal.pone.0229003 PMID:32160189

The University of Manchester Library. (2020). *My Research Essentials*. Retrieved from https://www.library.manchester.ac.uk/using-the-library/staff/research/my-research-essentials/

University of Nairobi. (2015). *UoN is now ISO 9001:2015 Certified*. Author.

University of Nairobi. (2020). *Student Admissions*. Retrieved from https://a.uonbi.ac.ke/admission

U.S. News & World Report. (2020). *2020 Best Global Universities in Africa*. Retrieved from https://www.usnews.com/education/best-global-universities/africa

Warren, E. (2016). Strengthening Research through Data Sharing. *The New England Journal of Medicine*, *375*(5), 401–403. doi:10.1056/NEJMp1607282 PMID:27518656

Woeber, C. (2017). *Towards best practice in research data management in the humanities.* Academic Press.

Johnson Mulongo Masinde is a Ph.D. Candidate at Central China Normal University. He has also authored several papers and has huge experiences in teaching Information Science courses in Kenyan Universities.

Daniel Wambiri Muthee (PhD) is a former head of the department, Library and Information Science at Kenyatta University and the current director of the career development and mentoring programs at the Institution. He has also authored several papers and supervised many students both at Ph.D., Masters, and Undergraduate.