
Research Data Handling by Researchers in the Selected Universities in Tanzania

Obadia Shadrack Buhomoli 
University of Dodoma
Email: obadiashadrack16@gmail.com

Paul S Muneja 
Information Studies Programme, University of Dar es Salaam
Email: muneja@udsm.ac.tz

Abstract

This study investigated the practice of research data handling in the selected higher learning institutions in Tanzania, with specific reference to the University of Dar es Salaam and University of Dodoma. The study assessed research data formats that are currently produced, where the research data are currently stored and whether researchers in the selected institutions use data management plan in their research activities. A study used both quantitative and qualitative approaches. Questionnaire was used to extract data from researchers while interview was used to obtain data from key informants who were categorized as potential managers responsible for research data management from the respective institutions. Results show that researchers have been producing various kinds of research data in various formats both digital and analogy data. Moreover, researchers handled these data using various approaches. This means that there were no systematic and well-guided methods of handling their research data making the data prone to get lost. Only a small number of researchers appeared to have a large amount of research data. Furthermore, only a few number of researchers indicated to have used research data management plan (RDM) in their research activities. This study recommends more collaboration among researchers, directorate of research and publication, ICT and library management to improve the research data handling in the universities.

Keywords: Research data, research data management, RDM, research data handling, higher learning institutions, Tanzania, open data.

<https://dx.doi.org/10.4314/udslj.v16i2.5>

Introduction

Research data handling is one among essential tasks in research data management (RDM). In order to promote and enhance the quality research data and transparency, data emanating from research can collected, stored, preserved and be re-used by other researchers. Several studies have highlighted several benefits of proper handling of research data. These include improving research data management activities, reducing duplication of research data collection and the ability for such data to be reused (Bangani & Moyo, 2019; Chigwada, Chiparausha, & Kasiroori, 2017; Deards, 2013; Gewin, 2015). Additionally, if data are handled properly using systems such as data repositories preferably open repositories, they could enhance the accountability of researchers and validation of research findings (Deards, 2013; Jaakkola, Mäkinen, & Eteläaho, 2014).



Universities and research institutions are mostly known for generating data from research because research is one of the three core objectives of universities and research expertise is heavily invested in these institutions. In view of this, generated data in these institutions are supposed to be handled properly in such a way that they can be used in the future in validating the research findings, to avoid duplication of effort and for cost cutting. However, in most African countries, it is not known how data are handled after the research project is over (Borgman, 2018). Currently, handling of research data is very crucial especially in this era where there is a proliferation of shared research output online. Therefore, validation and transparency of data is inevitable in promoting quality and impact of research. Aurban University (2019) stipulates that research data have to be handled with care and be made easily available when the need arises. However, if research data are not handled properly, there might be duplication of efforts resulting into wastage of time and money needed for re-collecting the same data (Karick, 2014)

Research data handling is one of the pre-requisites towards adoption of open data. Open data are known for increasing accountability, transparency, research visibility, and validation of research findings and protecting the integrity of the research findings (Bangani & Moyo, 2019; Zuiderwijk & Spiers, 2019). Several academic and research institutions are currently collecting, storing and preserving research data in open data repositories. Few examples of these according to <https://www.re3data.org> include data repositories at University of Sheffield, University of Cape Town, University of Pretoria and University of Essex. The same list can be traced from Open Access Directory of Data Repositories (OAD) and <https://www.fairsharing> (Tripathi, Shukla, & Sonker, 2017).

Handling of research data in Africa has been reported to be poorer compared to some other continents (Avuglah, 2016; Davies, Walker, Rubinstein, & Perini, 2019). However, the situation in sub-Saharan Africa is still at infant stage with exception of South Africa which is far better than the rest of the African continent in terms of research data management (Bangani & Moyo, 2019; Joseph, 2017; Ng'engo, 2018). The government of the United Republic of Tanzania (URT) has shown to have made some steps towards management of public data by establishing open data repository that is accessible through <https://www.opendata.go.tz>. This data repository contains various data sets generated by the government agencies including health, education, water and other sectors. The establishment of open data repository went hand in hand with formulation of open data policy where in 2016; the first draft of the policy was released. The establishment of data policy in the country may provide guidelines on handling of research data (The United Republic of Tanzania, 2016). Other initiatives that have been pioneered by the URT that may further support research data handling include establishment of the National ICT Broadband, which have stabilized internet accessibility, and establishment of institutions, which deal with research data. These are such as; National Bureau for Statistics (NBS), Commission for Science and Technology (COSTECH), e-Government Agency (eGA), Records and Archives Management Department (RAMD), Tanzania Information Services and other public institutions which are dealing with research data (Anatory, 2017; Mulamula, 2016; Shao & Saxena, 2019). These institutions do guide, manage and regulate the use of data in one way or the other (Shao & Saxena, 2019) For instance, NBS is responsible for coordinating and managing data contents in the government open data repository. eGA is responsible for maintaining the data repository and provides technical support. Furthermore, RAMD is responsible for classification and verification of datasets; (Shao & Saxena, 2019). Tanzania Communications Regulatory Authority (TCRA) is responsible for the provision of infrastructures to manage online data while Information Services

Research Data Handling by Researchers in the Selected Universities in Tanzania

(MAELEZO) is responsible for raising awareness on data handling and open data (Shao & Saxena, 2019; Tanzania Communications Regulatory Authority, 2020; The United Republic of Tanzania, 2016).

In academic sector, proper handling of research data plays a very significant role in academic and research activities. The proper handling of research data facilitates transparency and accountability in research processes as it allows validation of research findings (Bangani & Moyo, 2019; Chigwada, Chiparousa & Kasiroori, 2017; Deards, 2013). Despite the benefits that come with proper research data handling, the state of research data handling by universities in Tanzania is not known. It is not clear how the generated data are collected, stored, preserved, and are made accessible for later use. Despite the efforts and initiatives made by the URT together with other institutions, little is known on how higher learning institutions handle their research data. The study by Mushi, Van Deventer and Pienaar (2020) investigated about implementation of Research Data Management (RDM) services at the Library of the University of Dodoma, Worker and Excell (2018) studied about open data for climate change, Shao and Saxena (2019) focused on barriers to open government data initiatives in Tanzania while Buhomoli and Muneja (2020) concentrated on awareness of open data among researchers in Tanzania. Nevertheless, none of these studies focused on research data handling by academicians, and this is the gap that this study intended to fill. Thus, this study aimed at investigating research data handling in selected higher learning institutions in Tanzania with special focus on the University of Dar es Salaam and University of Dodoma. Specifically, the study aimed to assess research data format that are currently produced, to identify how research data are currently handled as well as to explore whether researchers in higher learning institutions use data management plan in their research processes. This study was guided by the following research questions:

- i. What type of data format is currently produced by academicians and researchers in the selected Universities?
- ii. How are research data currently handled in the selected Universities?
- iii. How does the academic and researchers in the selected Universities comply with research data management plan in their research processes?

Literature Review

Literature of this study is organized based on the objectives of the study. The plan of literature starts by discussing the general concept of research data handling, then it goes down to the specific objectives of the study.

Research Data Handling

Research data handling is one of the essential tasks in research data management (Chigwada *et al.*, 2017; Kennan, 2016). Proper handling of research data is a crucial aspect towards adoption of the open data best practices as opposed to poor handling that acts as an obstacle in adoption of open data (Apanasevic, 2018; European Research Council, 2018). Handling of research data as one of the key aspects of Research Data Management (RDM), aims at assessing the formats of research data that are produced by researchers, the amount of research data that are possessed by researchers, duration in which data is kept by the researcher, place at which data is stored as well as the implementation of Research Data Management Plan in the research processes (Apanasevic,

2018; Ng'engo, 2018; Patrick *et al.*, 2013). Research data are recognized as the valuable and important resources, which have to be taken with care. A study by Wyk (2018) conducted in South Africa revealed that most of the Universities in South Africa had devised mechanisms for handling their research data that made them reap the benefits associated with data. The same was observed in studies which were conducted in America, India, Wales (UK) and Austria which all showed efforts which were undertaken to strengthen data handling and management in their respective countries (Buys & Shaw, 2015; Kennan & Markauskaite, 2015; Tripathi *et al.*, 2017; Van den Eynden *et al.*, 2016).

Universities are currently using various ways of handling research data such as establishment of institutional data repositories where researchers deposit their data for the purpose of storing, preserving and sharing them with other researchers (Bangani & Moyo, 2019; Chigwada *et al.*, 2017; Tuyl & Michalek, 2015). Despite the fact that some of the Universities, do not have data repositories, they use policies or guidelines which mandate their researchers where to store data and at what period (Piennar, 2010; Tripathi *et al.*, 2017; Wyk, 2018). However, even those Universities which have data repositories, they lack guidelines which direct researchers where to store research data and the time to do so (Ng'engo, 2018; Tuyl & Michalek, 2015). Scholars have shown that, researchers whose institutions do not have data repositories are using various storage facilities in storing their research data. The tools which are commonly used are personal computer saver, online storage, personal devices and others (Kennan, 2016; Ng'engo, 2018). However, more challenges have been noted especially when a researcher has a large amount of data which cannot easily be accommodated with their storage devices.

Besides that, studies which were conducted by Kennan and Markauskaite (2015) and Chigwada *et al.*, (2017) on practices of research data management in Wales and Australia and Zimbabwe respectively, showed that researchers in Wales and Zimbabwe were responsible for handling and management of their research data. In addition, it was upon the researchers' decision to decide on whether to store them for the future use or to discard them. Studies conducted by Elsayed and Saleh (2018) in Arab countries and Ng'engo (2018) in Kenya also revealed the same results. Moreover, Kennan and Markauskaite (2015) argue that it is upon the researchers' preference to decide and choose places for storing their research data for example: researchers in Australia preferred to store their research data in Compact Disks, USB drives computer hard disk, external disk, and very few respondents reported to store their data on the cloud. This indicates that, researchers were using various ways for the storage of their research data in higher learning institutions.

Format of Research Data

Formats by which research data are produced indicate how the research data are handled in the respective organization (Whitmire, Boock, & Sutton, 2015). Existence of multiple data formats in one institution may imply lack or no guideline on data handling (Ng'engo, 2018; Whitmire *et al.*, 2015). Policies and guidelines help in the implementation of proper data handling by researchers from production and preserving them in the format that have been assessed and accepted by institutions. However, some data exist in analogy form, the format that is non-machine readable which may cost an institution which wants to convert them into digital formats especially those institutions that want to adopt research data management initiatives (Kennan & Markauskaite, 2015; Ng'engo, 2018; Tuyl & Michalek, 2015). It is also true to Tuyl and Michalek, (2015) who

Research Data Handling by Researchers in the Selected Universities in Tanzania

argue that conversion of data from one format to another always requires application of hardware and software to implement and some software may require the proprietary software which are costly. Scholars such as Chigwada *et al.*, (2017); Chiware and Becker, (2018) and Wyk, (2018) recommend file formats which are produced by researchers to be in format that can easily be contextualized. This means the format that does not necessarily require proprietary software as well the format that does not require any technical assistance in order for it to be used. In other words, they recommend the use of open file formats that are using open source software (OSS) which are easily manageable.

The study by Chigwada *et al.*, (2017) presented that researchers in Zimbabwe research institutions were keeping their research data in various file formats such as spreadsheets, video, text documents, graphics, audio, and structured text. A study conducted by Bangani and Moyo (2019) also asserted that research data that were produced by researchers in South African Universities were mostly in office formats (64%), statistical format such as SPSS, GIS, FITS (36%), web-based research data set (19%), image (15%) and archive data such as ZIP, RAR, JAR (0.8%). Moreover, majority of the studies on research data management and data handling among researchers found that most of the research data were produced in office formats and fewer of them in the archive data formats (Elsayed & Saleh, 2018; Van den Eynden, Corti, Woollard, Bishop, & Horton, 2011; Van den Eynden *et al.*, 2016). Ng'engo (2018) argues that knowing the file formats produced by researchers is very important since it indicates how the researchers are handling their research data and whether the formats would favour future use which is the most important and indication of the value of research data.

Research Data Management Plan

Research Data Management Plan (RDMP) is very important in assessing the current handling of research data. Ng'engo (2018) defines the plan of managing research data as the document that shows how data would be collected, analysed, where should they be stored, at what duration, who would have the access to them, and how will they use the accessed data. RDMP can be treated as the roadmap of managing and handling research data. Some of the institutions require their researchers to provide the data management plan before going to the field. RDMP is used to control the handling and managing of research data, as well as acting as an important tool for data quality assurance (Data curation centre, 2004; UK data Archives, 2017).

The study conducted by Tuyl and Michalek (2015) found that most of the researchers at Carnegie Mellon University were using RDMP in their research processes especially for the public funded research projects as opposed to Kennan and Markauskaite (2015) and Ng'engo (2018) who found that researchers in Wales, Australia and Kenya were not using RDMP in their research processes. Perhaps the difference of using or not using RDMP could be attributed to by several factors including policies and guidelines, ICT infrastructure, knowledge and so on.

Literature suggests that one of the most important step in the adoption and usage of open data is to know the current practices of handling research data. Literature has shown that researchers were producing research data in both digital formats and analogy formats (Bangani & Moyo, 2019; Chigwada *et al.*, 2017; Elsayed & Saleh, 2018). Research data are mostly stored by researchers through their personal storage systems (Kennan, 2016), although some researchers have started depositing their data into data repositories (Koopman & Jager, 2016a). Moreover, the usage of RDMP appeared to vary across countries and researchers. Knowledge on current handling of research data by institutions has a direct link with their readiness in adoption open data.

Methodology

This study was conducted in the University of Dar es Salaam and University of Dodoma in Tanzania. Both sites at which the study was conducted were selected because they are the largest public universities in the country with a lot of researchers, reliable ICT Infrastructures and because of their potentiality in the use and reuse of research data. Authors chose public Universities because they are funded by the government using public money where it is believed that the proper handling of research data will eliminate wastage of public funds caused by duplication of research data collection. A cross-sectional survey design was employed in the study where both quantitative and qualitative data collection methods were employed. Quantifiable data were collected using quantitative methods while qualitative data were collected using qualitative methods. The study population included researchers and potential research data managers. This group of the respondents was used because it was believed that it would give the research insights on research data handling considering that this topic was a new phenomenon in Tanzania during the time of this survey. The sample size was calculated using Yamane's (1967) formula. The sampling frame was determined from the facts and figures' documents of UDOM and UDSM (University of Dar es Salaam, 2017; University of Dodoma, 2018). At the time when this study was being conducted, UDOM had 782 academic staff and UDSM had the 1164 academic staff.

The confidence interval employed by the study was 90%, and sample size was calculated and derived from the Yamane's formula and found to be 95. This meant that the number of respondents to participate in this study was supposed to be not less than 95. Taking into account the fact that not all individuals would return the questionnaires distributed to them, the number of individuals to participate in this study was set at 200. The distribution of the sample between the two universities was calculated based on the number of researchers that each institution had, thus contributing to the total study population. One hundred and forty-seven (147) questionnaires were returned which made a response rate of 73.5%. The distribution of questionnaires for each of the participating universities was based on the available numbers of the academic staff. Based on this criterion, 84 questionnaires were distributed to UDOM academic staff and 116 questionnaires were distributed to UDSM academics staff. More details are indicated in Table 1.

Table 1: Sample size distribution between the selected higher learning institutions

	Population	Percentage of the total population	Sample Size
UDOM	782	42%	84
UDSM	1164	58%	116
Total	1846	100%	200

Source: Field Data (2020)

Both probability and non-probability techniques were employed in this study. Stratified sampling was applied to obtain respondents who filled in the questionnaires. Purposive sampling was applied to obtain potential research data managers who participated in key informants' interviews. Research data managers are administrators or directors overseeing all workflows related to [Research Data Handling by Researchers in the Selected Universities in Tanzania](#)

research data (Avuglah, 2016; Ng’engo, 2018). In this study, research data managers included all directors who were the potential overseers of research data. University top leaders/managers were chosen based on their roles in handling research data where six potential managers from each institution were chosen including Directors of Library Services, Directors of Research and Publication, Directors of Postgraduate Studies, Directors of Quality Assurance, and Directors of ICT. To maintain anonymity and research ethics, names of the respondents together with their affiliations, were hidden. Instead, respondents were given codes starting with R1 up to R12. The quantitative data collected from the survey were analysed quantitatively using Statistical Product for Service Solution (SPSS) while qualitative data were analysed qualitatively using content analysis.

Presentation of the Findings and Discussion

This section presents the findings together with the discussion of each finding. The section is organized according to the research questions which guided this study.

Respondents' Demographic Characteristics

The demographic characteristics of the respondents were as follows: female respondents were 58 (39.5%) lower than males who were 89 (60.5%). This finding suggests that the number of female researchers in academic institutions in most of the African countries is still low (World Health Organization, 2015). The age of the respondents were as follows: the respondents with ages between 36 and 45 years were 57 (38.8%), those aged between 26 and 35 were 44 (29.9%), those aged above 55 years were 20 (13.6 %.), respondents aged 46 and 55 were 20 (13.6%), and those with the age of 18 and 25 were 6;(4.1%). The education level of the respondents was as follows: PhD degree 57 (38.8%), master’s degree 84(57.1%), and (6 respondents (4.1%) had bachelor degree. These findings align with the study by Kimaryo (2016) who found higher number of academic members in higher learning institutions with master’s degree than other education levels. This level of education was sufficient to provide information on research data handling in their institutions. Therefore, this study sought to capture the said demographic features and see whether they had influence on research data handling.

Publication Experience

This study sought to understand the respondents’ experience in publication in order to ascertain researchers’ current practices in managing their research data. Table 2 reveals that 49 (33.3%) of the respondents had experience of 3-5, while 45 (30.6%) had experience of 6-8 years, 35 (23.8%) had experience of more than 8 years, 18 (12.2%) had experience of less than three years’ in conducting research.

Table 2: Experience in doing Research (n = 147)

INST.	Less than 3 years		3-5 years		6-8 years		Above 8 years		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq	%	Freq.	%
UDSM	8	5.4	26	17.7	28	19.0	22	15.0	84	57.1
UDOM	10	6.8	23	15.6	17	11.6	13	8.8	63	42.9



TOTAL 18 12.2 49 33.3 45 30.6 35 23.8 147 100

Source: Field Data (2020)

These results imply that most of the respondents had sufficient experience in doing research. Knowing research experiences of the respondents in doing research was very important since such an experience was likely to have a profound influence on researchers' decisions regarding handling of their research data (Kassen, 2018).

File Formats Produced in the Course of Research

Results in Figure 1, show that respondents were generating their data in a variety of formats. The findings indicate that; 73 (49.7%) were generating their files in print format, 115 (78.2%) were generating research in text documents (.pdf, .docx, .rtf, .doc, .log., .txt, etc.). Again, 82 (55.8%) were generating their research data in spreadsheet file formats (csv, .wks, ods, .xls, etc.). Moreover, 30 (20.4%) were generating their research data in a structured text/ Web (xml, Html, json, xhtml, etc.). The 50 respondents (34%) were generating their file formats in graphics/ Images (.png, .bmp, .jpg, .tiff, svg, .gif, etc.). It was also found that 25(17%) were generating the research in audio file formats (ogg, .mp3, .wav, aiff, etc.) 16 (10.9%) were generating their file formats in video/Film files (mp4, mpeg, wmv, mov, avi, etc.), 14 (9.5%) were generating the file formats in computer aided design (CAD), 80 (54.4%) statistical data (.sas, .spss, .sav, etc.) while only 5 (3.4%) were generating their findings in other file formats.

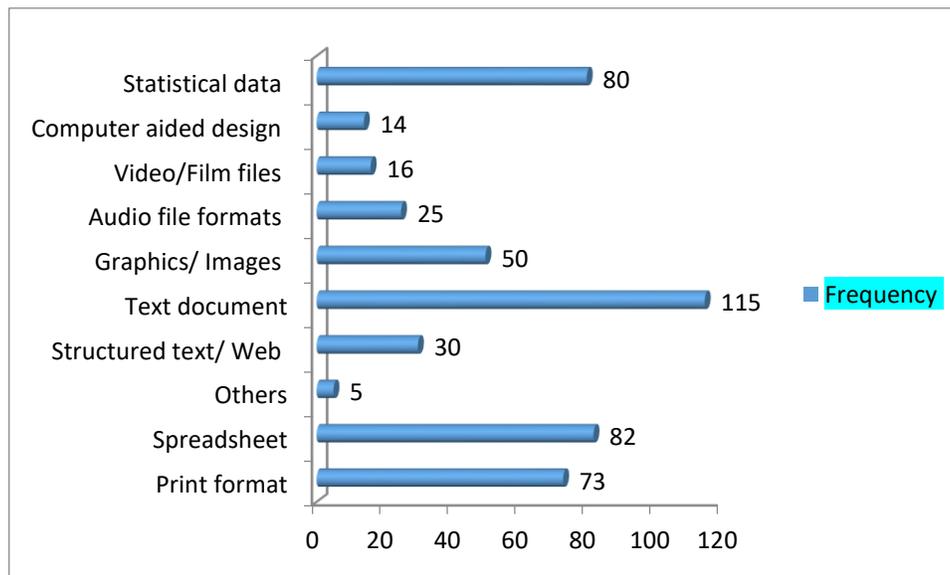


Figure 1: Research data formats in the area of study (n=147)

The findings show that when this study was being conducted, researchers were generating data in both digital and analogy format. Managing data in a variety of format may have challenges in curation, security and their accessibility. This concern was expressed by Chigwada *et al.*, (2017), Johnson, Chiareli and Kaye (2016) and Kennan and Markauskaite (2015) who cautioned about managing research data in a variety of formats. The findings of this study also cement the study

by Bangani and Moyo, (2019) who emphasized on the importance of guidelines for the generation of research data. The findings of this study imply that, there were no guidelines on the type of data format that the researchers would follow.

Research Data Handling

To answer a question on how research data were being handled, researchers investigated the following variables: research data storage, handling of research data after the research project was over, duration of the keeping research data and the amount of research data that were possessed by the researchers.

Research Data Storage Devices mostly used by the Respondents

Respondents were asked to indicate the type of storage devices they were mostly using in storing their data. The findings show that the respondents were using various devices for the storing their research data as illustrated in Figure 2.

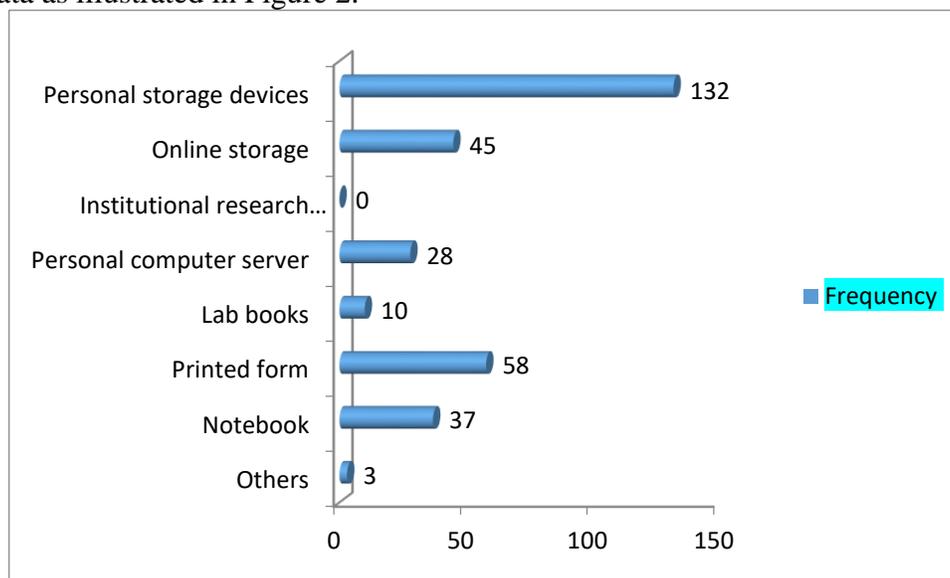


Figure 2: Storage devices mostly used by the Respondents (n=147)
Source: Field Data (2020)

As presented in Figure 2, the findings indicate that majority of the respondents were storing their data in personal storage devices such as flash drives USB, computer hard drive, DVDs, CDs, 132 (89.8%). It was noted that few respondents were using online storage such clouds etc. 45(30.6%). Those who were using personal computer servers were 28 (19.0%). Again, those who were using notebooks were 37 (25.2%), those who were using printed forms were 58(38.5%), while those who

were using other means of data storage were 3(2%) and research data repository was not mentioned by any one.

As presented above, the findings show that there was no single way that was being used by researchers in storing their research data. As the findings reveal, researchers were storing their research data depending on conveniently available means at a time. This implies that, researchers lacked a proper designed database (physical or online) where they could store and re-use their research data when needed in the future. These findings correspond with those by Avuglah (2016), Bangani and Moyo (2019) and Mushi *et al.*, (2020) who also found lack of proper designed databases for data storage.

Although Figure 2 indicates that the most commonly used were personal storage devices such as HDD, SSD, CDs, DVDs, and flash drives etc, these devices could easily lose data due to their physical delicateness, less secured as they could be attacked by the computer malware, thus leading to the data loss (Ng'engo, 2018). On the contrary, Johnson *et al.*, (2016) found that most of the researchers liked cloud storage for storing their research data thus differing with the findings by Kennan and Markauskaite (2015) who found that most of the researchers in Wales and Australia liked using unsecured devices for storing their research data. Again, these findings show that there were no any guidelines on how research data could be treated as researchers were preserving their data according to what they perceived as right to them. These findings relate with those by Avuglah (2016) in Ghana, who noted lack of coordination on how research data could be treated. Additionally, interviews with the directors responsible for Research and Publications and those responsible for Library Services, showed that the issue of handling of research data was left to the individual researchers to decide. These findings contrast with those by Ng'engo, (2018) who indicated that research data handling in most of the research institutions in Kenya was directed to departments responsible for research data management in their respective departments.

Research Data Handling After the Research Project

Respondents were also requested to indicate the way they were treating research data after the research data projects. Figure 3 shows the results obtained from the field.

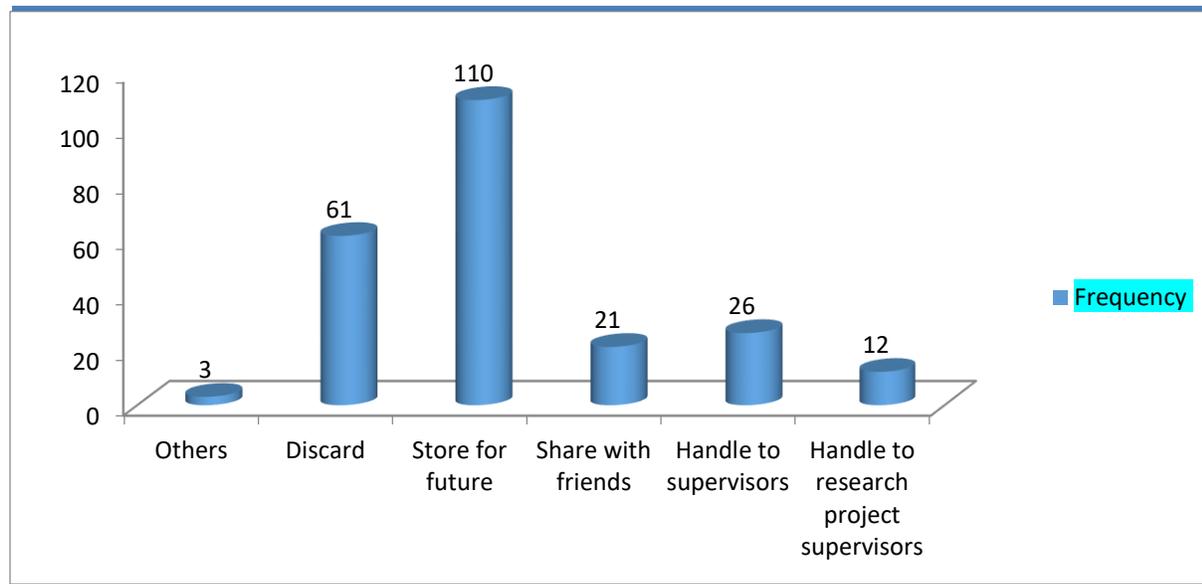


Figure 3: Treatment of research data after the research project (n=147)
Source: Field Data (2020)

The findings in figure 3 indicate that, most of the respondents revealed that, they were storing their data for the future use 110 (71.8%). Other respondents 61(41.5%) said that they used to discard their research data, (26 (17.7%) they used to handle their research data to the project supervisors, 21 (14.3%) they used to share their data with friends and 12 (8.2%) said that they used to handle their research data to their respective research offices.

Similarly, interviews with R5, R11, R3 and R9 revealed that matters with research data handling were left to individual researchers. These findings contrast with those by Ng'engo (2018), where research data handling was found to be done at the departmental level where departments were in-charge of all the activities regarding research data management. In this study, key informant interviews also indicated that researchers owned their research data and had to decide whether to keep them for future use or not. Apart from that, researchers were responsible for choosing places to store their data if a need to do so arose. According to Koopman and Jager (2016b) keeping data in this way could be a primary inhibitor of the adoption of OD initiatives. When asked about the current data storage place, R5 responded as follows:

Our directorate is much concerned with research outputs. Research data handling is left to the individual researcher”.

Almost the same response was given by R11 who said:

Research data handling is the task of the individual researcher; it is not our task as the Directorate. We are concerned with research outputs”.

As such, Key informants interview results revealed that the higher learning institutions selected for this study were only responsible for research outputs not the data used to produce them. This was so even for data collected in the public funded research projects. Due to this, it was noted that valuable and expensive datasets were discarded and got lost. This is to say, because, research data

were handled individually, for the case where researchers left or disengaged from the institutions, the research data they had would usually get lost or left unattended. These instances would lead to the duplication of research efforts and financial loss. In the same vein, Chigwada *et al.* (2017) also noted lack of coordination in research data handling in Zimbabwe research institutions.

Duration for Keeping Research Data

Respondents were asked how long were they keeping the research data after their research data was done. The findings revealed that the respondents had a different time range for keeping their research data. This is as indicated in Table 3 below.

Table 3: Duration for Keeping Research Data (n=147)

INSTITUTION	UDSM		UDOM		TOTAL	
	Freq.	%	Freq.	%	Freq.	%
Never stored my research data	6	4.1	2	1.4	8	5.4
Less than a year	12	8.2	9	6.1	21	14.3
1–3 years	25	17.0	24	16.3	49	33.3
4–6 years	18	12.2	15	10.2	33	22.4
More than 6 years	23	15.6	13	8.8	36	24.5
Total	84	57.1	63	42.9	147	100

Source: Field Data (2020)

Results in Table 3 indicate that 21 respondents (14.3%) used to keep their data for less than a year, 49 (33.3%) of the respondents indicated that they used to keep their research data for 1-3 years. Again, 33 respondents (22.4%) indicated that they used to keep their data for 4-6 years and 36 (24.55%) indicated that they used to keep their research data for more than 6 years. Although researchers claimed to have been storing their research data for future use, surprisingly, very few of them indicated to have been storing their research data for six years and above. This practice could be attributed to by the fact that there was no research data repository that was being used to archive and preserve research data in the surveyed institutions. Having different time ranges of keeping research data among researchers could also mean lack of guidelines and coordination regarding the actual time of keeping data (Chigwada *et al.*, 2017; Ng'engo, 2018).

Amount of Research Data produced After Research

It was found that the researchers in the selected higher learning institutions were generating and storing different amounts of research data as shown in Table 4.

Table Error! No text of specified style in document.: Amount of Research data generated by Respondents (n=147)

INSTITUTIONS	UDSM		UDOM		TOTAL	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
None	5	3.4	4	2.7	11	7.5
Less than 1GB	36	24.5	29	19.7	65	44.2
1 to 10GB	28	19	24	16.3	52	35.4

Research Data Handling by Researchers in the Selected Universities in Tanzania

10 GB to 1TB	8	5.4	6	4.1	14	9.5
1 to 10TB	7	4.8	0	0.0	7	4.8
More than 10TB	0	0.0	0	0.0	0	0.0
Total	84	57.1	63	42.9	147	100

Source: Field data (2020)

The findings as shown in Table 4 indicate that the amount of research data that were possessed by researchers in the selected institutions were as follows: 65 (44.2%) were possessing less than 1 GB research data, 52 (35.4%) had possessed 1 to 10 GB of research data. Again, (14; 9.5%), possessed 10 GB to 1 TB of research data, 7 (4.8%) were possessing 1 to 10TB of research data and (11; 7.5%) were not possessing any research data when this survey was conducted. The findings also show that, the number of the respondents who appeared to possess a large number of research data sets in the surveyed institutions was small implying that the generated research data could easily be handled by the research data repositories if established. Those few individuals who had large amount of research data sets would better call for the establishments of the research data repositories as it might reach a point where the capacity of collected data could not be easily handled by an individual researcher thus, a need for a large storage capacity (Kennan & Markauskaite, 2015; Whitmire *et al.*, 2015). Tuyl & Michalek, (2015) Mary Anne Kennan & Markauskaite (2015) share the same results.

Compliance and Non-Compliance with the use of Research Data Management Plan (RDMP)

Research data management plan as one of the important elements of research data management was examined in this study for knowing the current practices in handling of research data in the selected higher learning institutions. The results of these studies are as portrayed in Table 5.

Table 5: Compliance with of the use of RDMP in research processes (n=147)

RDMP	UDSM		UDOM		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Use RDMP	32	21.7	18	12.3	50	34
Don't use RDMP	52	35.4	45	30.6	97	66
TOTAL	84	57.1	63	42.9	147	100

Source: Field data (2020)

The results in Table 5 show that only (50 (34.0%) of the respondents in the selected higher learning institutions were using Research Data Management Plan in their research processes and the remaining 97 (66.0%) were not.

This study also found that RDMP was not being used and was not mandatory to be used in the research processes in the selected higher learning institutions. These results match with those by Kennan and Markauskaite (2015) and Ng'engo (2018) who found that, researchers in Wales, Australia and Kenya were not using RDMP in their research processes. Failure to use RDMP in research process could affect the research data life cycle as there might be poor guidance of the collection, handling, storage and may affect data re-use. The general understanding shows that failure to use RDMP could lead to poor production, management, and poor storage of research

data because of lack of guidance. As such, the low usage of such a plan recorded at the universities under study is likely to point to poor observation of some research data management principles despite the indications that things were getting done. For instance, responding to a question on RDMP, R11 stated as follows:

Here in our institution, it is not mandatory for our researchers to use RDMP in their research processes. The decision to use of RDMP is upon the researcher himself/herself.

This appears to explain the low usage of RDMP in the selected higher learning institutions even in research projects operating under public funding. These findings are different from those by Tuyl and Michalek (2015) who found that researchers in Carnegie Mellon University were mostly using RDMP in research projects operating under public funding. On the whole, the low usage of RDMP in the two institutions of higher learning involved in this study is an indication that adoption and use of OD are likely to be faced by difficulties since the quality of research data is probably compromised (Avuglah, 2016; ERC,2018; Ng’engo, 2018).

Respondents who said they were not using RDMP in their research process were then asked to indicate the reasons for not using it. This question was also important as it intended to bring to light the existing practices of management research data in the selected higher learning institutions. To respond to this question, respondents were required to tick all that apply. The results on this are shown in Figure 2.

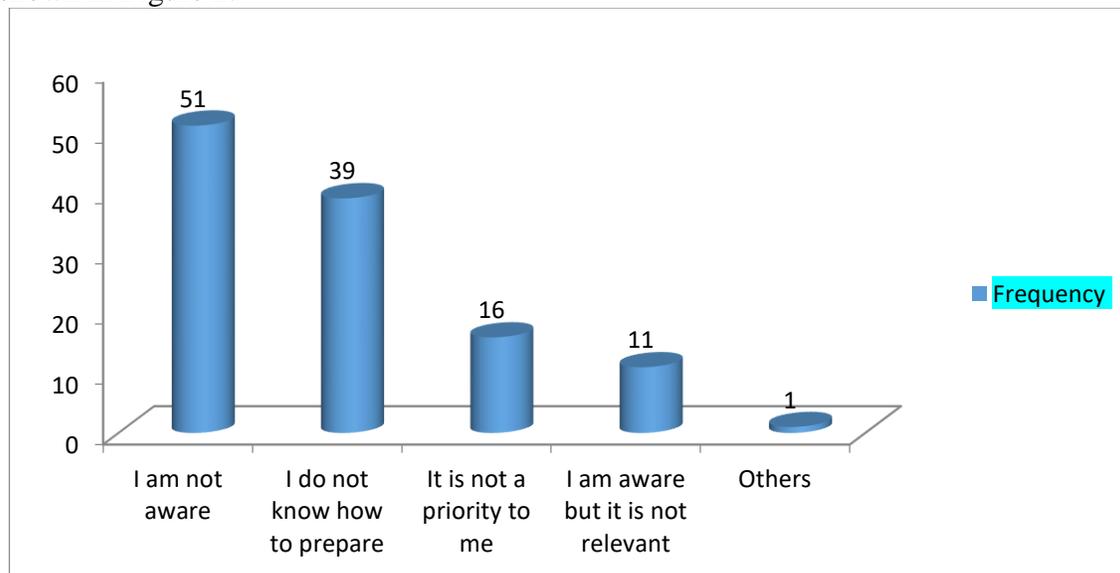


Figure 4: Non-compliance with the use of RDMP (n=97)
Source: Field Data (2020)

It was found that the respondents had different reasons for not using RDMP in their research processes. These reasons were such as being unaware of the plan 51 (52.6%), inability to prepare RDMP 39 (40.2%), the plan not being a priority 16 (16.5%), and irrelevance of the plan to studies 11(11.3%). These results are in some way similar to those reported by Tuyl and Michalek (2015) where was revealed that most researchers were not using RDMP because they found it irrelevant to their research areas or simply unimportant. This further shows that units responsible for research

data management have a very huge task to perform regarding making their institutions ready for OD, including promoting the usage of RDMP in the research processes. Lastly, the findings of this study imply that there is a need for revising HLIs research policies and guidelines to accommodate issues related to research data handling.

Conclusion, Recommendations and Implications

This study concludes that; research data were not properly being handled in the area of this study. It was found that data handling was left in the hands of individual researchers. This is supported by the fact that most researchers were using personal devices for storing their research data. It was found that there were no any established guidelines on the type of research data file formats to be produced by researcher in the respective institutions. Again, it was found that research data were owned by individual researchers not by their respective institutions thus making them decide whether to keep or discard after the end of the research project. Furthermore, there was lack of established data repository for archiving and preserving research data. Lack of proper data handling was echoed by the unavailability of RDMP even for the publicly funded research projects. This study proposes the establishments of legal and institutional frameworks to guide the issue of research data handling and research data management at large from the data capture stage to data disposition. This study also recommends that, the units responsible for research and publications in the respective institutions, should not only concentrate their efforts on the research outputs, but also they should direct their efforts to the management of research data. In addition, data literacy training should continuously be conducted to researchers in order to increase their knowledge on data handling. This study therefore recommends more collaboration between researchers, directorate of research and publication, ICT and library to improve the research data handling.

References

- Anatory, J. (2017). ICT for fostering industrilization and socio-economic deveopment in Tanzania. *Sufficiency Economy Philosophy (SEP) and Sustainable Development Goals (SDG)*. Retrieved from <http://www.erb.go.tz/index.php/about-us/55-annual-eng-day-2017>
- Apanasevic, T. (2018). *Opportunities and challenges of mobile payment services: The perspective of service providers* (Doctoral dissertation, Kungliga Tekniska högskolan).
- Aurban University. (2019). *Open data and academia: What role does academia play in helping to open data to solve world hunger*.
- Avuglah, B. K. (2016). *Developing and implementing plan for research data management at the University of Ghana*. Master's dissertation, University of Pretoria, South Africa.
- Bangani, S., & Moyo, M. (2019). Data sharing practices among researchers at South African Universities. *Data Science Journal*, 18(1).
- Borgman, C. L. (2018). *Open data, grey data, and stewardship: Universities at the privacy frontier* *Berkeley Technology Law Journal*, 33 (2), 365-412.
- Buhomoli, O. S., & Muneja, P. S. (2020). Awareness of open data among researchers in selected public Universities in Tanzania. *University of Dar es Salaam Library Journal*, 15(1), 173–185.
- Buys, C. M., & Shaw, P. L. (2015). Data management practices across an institution : Survey and report data. *Journal of Librarianship and Scholarly Communication*, 3(2), 1–24.

- Chigwada, J., Chiparausha, B., & Kasiroori, J. (2017). Research data management in research institutions in Zimbabwe. *Data Science Journal*, 16(31), 1-9
- Chiware, E. R., & Becker, D. A. (2018). Research data management services in southern Africa: a readiness survey of academic and research libraries. *African Journal of Library Archives and Information Science*, 28(1), 1-16.
- Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. (Eds.). (2019). *The state of open data*. Cape Town: African Minds and International Development Research Centre.
- Deards, K. (2013). An open data manifesto. In Council on library and information resources (Ed.), *Research data management principles, practices, and prospects*. University of Texas.
- Elsayed, A. M., & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA Journal*, 44(4), 281–299.
- European Research Council. (2018). *Open research data and data management plans information for ERC grantees*. (February), 1–15.
- Gewin, V. (2015). An open mind on open data. *Nature*, 529, 117–119.
- Jaakkola, H., Mäkinen, T., & Eteläaho, A. (2014). Open data - Opportunities and challenges. *ACM International Conference Proceeding Series*, 883, 25–39. New York, New York, USA: Association for computing machinery. <https://doi.org/10.1145/2659532.2659594>
- Johnson, R., Chiarelli, A., & Kaye, J. (2016). Research data assessment support: Findings of the 2016 data assessment framework (DAF) surveys. *Jisc*, (July 2017). Retrieved from <https://doi.org/10.5281/zenodo.177856>.
- Joseph, H. (2017). *What's Up with Open Research Data. Colorado Open Scholars. 1. Retrived from, https://digitalcommons.du.edu/co_Open_Scholarship/1*
- Karick, A. (2014). *Research data management: guidelines and planning for researchers*. Retrieved from https://www.swinburne.edu.au/media/swinburne/duau/research/%0Adocs/pdfs/Research_Data_Management_Guidelines.pdf.
- Kassen, M. (2018). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. *Aslib Journal of Information Management*.70(5)
- Kennan, M.A. (2016). Data management: knowledge and skills required in research, scientific and technical organizations. Retrieved from <http://library.ifla.org/1466/1/221-kennan-en.pdf>
- Kennan, Mary Anne, & Markauskaite, L. (2015). Research data management practices: A snapshot in time. *International Journal of Digital Curation*, 10(2), 69–95.
- Kimaryo, C. (2016). *Knowledge management initiatives in selected research institutions in Tanzania. PhD thesis*, University of Dar es Salaam, Tanzania.
- Koopman, M. M., & Jager, K. de. (2016a). Archiving South African digital research data : How ready are we ? The South African context. *South African Journal of Science*, 112(7), 1–7.
- Koopman, M. M., & Jager, K. de. (2016b). Archiving South African digital research data : How ready are we ? The South African context. *South African Journal of Science*, 112(7), 1–7.
- Mulamula, G. (2016). ICT and innovation: Tanzania National experience and challenges. *TANZIC Conference*, 1–15. Dar es salaam. Retrieved from <https://tanzict.com/2016.pdf>
- Mushi, G., Deventer, M. Van, & Pienaar, H. (2020). Identifying and implementing relevant research data management services for the library at the University of Dodoma. *Data Science Journal*, 19(1), 1–9.
- Ng'engo, E. G. (2018). *Research data management in Kenya's agricultural research institutes*. PhD thesis, University of Kwazulu Natal, South Africa.

- Pienaar, H. (2010). Survey of research data management practices at the University of Pretoria, South Africa: October 2009–March 2010.
- Shao, D. D., & Saxena, S. (2019). Barriers to Open Government Data (OGD) initiative in Tanzania : Stakeholders ’ perspectives. *Growth and Change*, 50, 470–485.
- Tanzania Communications Regulatory Authority. (2020). Delivering Beyond Promises. *Regulator*, (1).
- The United Republic of Tanzania. (2016). *United Republic of Tanzania Open Data Policy Draft*.
- Tripathi, M., Shukla, A., & Sonkar, S. K. (2017). Research data management practices in university libraries: a study. *DESIDOC Journal of Library & Information Technology*, 37(6), 417.
- UK data Archives. (2017). *Create and manage data: Quality assurance*. Retrieved from <http://www.data-archive.ac.uk/create-manage/quality>.
- University of Dar es Salaam. (2017). *Facts and figures from 2011/12 to 2016/17*. Dar es Salaam of Dar es Salaam.
- University of Dodoma. (2018). *Institutional Repository Policy*. University of Dodoma.
- Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). *Managing and Sharing Data: Best Practices for Researchers* (3rd ed.). Essex: UK Data Archive. Retrieved from <http://data-archive.ac.uk/media/managingsharing.pdf>.
- Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., ... Corti, L. (2016). *Towards open research: Practices, experiences, barriers and opportunities*.
- Van Tuyl, S., & Michalek, G. (2015). Assessing research data management practices of faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication*, 3(3). <https://doi.org/10.7710/2162-3309.1258>
- Van Wyk, B. J. (2018). *The relationship between research data management and virtual research environments*. PhD thesis, University of Pretoria, South Africa.
- Whitmire, A. L., Boock, M., & Sutton, S. C. (2015). Variability in academic research data management practices : implications for data services development from a faculty survey. *Electronic Library and Information Systems*, 49, 382–407. <https://doi.org/10.1108/PROG-02-2015-0017>
- Worker, J., & Excell, C. (2018). Climate change open data for sustainable development: Case studies from Tanzania and Siera Leone. *Global Partership for Sustaiable Development Data*.
- World Health Organization. (2015). *Africa’s women in science*.
- Yamane, T. (1967). *Statistics, an introductory analysis*, (2nd ed). New York: Harper and Row.
- Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data : A case study of motivations in astrophysics. *International Journal of Information Management*, 49(November 2018), 228–241. <https://doi.org/10.1016/j.ijinfomgt.2019.05.024>