

Data preservation practices for enhancing agricultural research data usage among agricultural researchers in Tanzania

Journal of Librarianship and
Information Science
1–13

© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09610006221138110
journals.sagepub.com/home/lis



Nolasko Victory Mwinami 

Sokoine University of Agriculture, United Republic of Tanzania

Frankwell W. Dulle

Sokoine University of Agriculture, United Republic of Tanzania

Wulystan Pius Mtega

Sokoine University of Agriculture, United Republic of Tanzania

Abstract

The objective of this study was to investigate the role of research data preservation for enhanced data usage among agricultural researchers in Tanzania. Specifically, the study aimed to examine the data preservation methods used by agriculture researchers, find out how long agriculture researchers preserve their agriculture research data, and determine factors that influence agriculture researchers on their choice of data preservation methods for use. The study employed a cross-sectional research design. The study employed both qualitative and quantitative approaches. A survey was conducted to collect data in 11 research institutions. A simple random sampling technique was used to select 204 respondents from the study area while purposive sampling techniques were used to select 11 agriculture research institutions including 10 Tanzanian Agricultural Research Institution (TARI) centers, and Sokoine University of Agriculture (SUA). Also, 12 respondents were selected purposively for an in-depth interview as key informants. The study adopted Data Curation Centre (DCC) Lifecycle Model to explain data preservation process. Findings indicated that a majority of more than 90% of researchers preferred to preserve their data using different storage devices such as field notebooks, computers, and institutional libraries. Moreover, findings indicated that about 74% of agricultural researchers preferred to preserve their data for more than 6 years after the end of the project. Findings also indicated factors that influence researchers in the choice of data preservation methods to be easy to reach, cost-effective storage devices, support to use the devices, adequate infrastructure for data preservation, and reliable power supply. It can be concluded that there is yet a great role of research data preservation in enhancing data usage among researchers in Tanzania. It is recommended that the government should establish an agricultural research data bank to guarantee permanent availability of data at all times when needed.

Keywords

Agricultural research data, data preservation, data reuse, data sharing, data storage

Introduction

Preservation of agricultural research data by research institutions is vital for enhanced sharing, reuse, and development of the agriculture sector. However, this depends on the choice of preservation method used. According to Kirub (2016), it is the primary role of Agriculture research institutes to collect, generate, store, and use these data. However, such data are faced with a threat of being lost at

the end of the project. Globally, it has been advocated that proper preservation of such data can guarantee easy access, browsing, consultation, and future use (Tripathi et al., 2017), and also can facilitate technology transfer, and

Corresponding author:

Nolasko Victory Mwinami, Sokoine University of Agriculture, P. O. Box 3158, Morogoro, 67112, Tanzania.
Email: mwinaminolasko@yahoo.com

innovation in agriculture (Ng'eno and Mutula, 2018a), creation of opportunities for access, sharing, and reuse by other researchers beyond a given geographical area (Brouder et al., 2019; Dileepkumar, 2014; Eckes et al., 2017; Granda and Blasczyk, 2011; Kirub, 2016; Zhao and Wang, 2015). The choice of the best preservation method is a pre-requisites for its success (Ng'eno and Mutula, 2018a). Data preservation practice has attracted more attention from research institutes, researchers, funding institutions, and international development agencies (Adrian et al., 2018; Kirub, 2016).

Some funding agencies and publisher advocate the deposition of data on open access repositories to facilitate easy access, and use/reuse by others (Tripathi et al., 2017). Likewise, the government of Tanzania has taken several initiatives to preserve different types of data. The establishment of electronic platforms at the National Bureau of Statistics (NBS), and Tanzania Meteorological Agency (TMA) facilitated the preservation and dissemination of scientific, baseline survey, and agriculture administrative data (Agrawal, 2017; Bhatia et al., 2016; TMA, 2020; URT, 2017). While these initiatives have been successful, there are limited initiatives that have focused on preserving agricultural research data in agriculture research institutions in Tanzania. Moreover, the SUA research regulation and guidelines of 2019 and TARI Act of 2016 both emphasize the preservation of research data and other research materials but the documents are silent on methods of data preservation for ensuring long-term data access, sharing, and future reuse (DPRTC, 2019; URT, 2016). Despite previous literature available on data preservation in the study area (Agrawal, 2017; Chidi, 2017; FAO, 2017; Katabalwa et al., 2021; Kijazi, 2018; Mboera, 2015; Mushi et al., 2020; NBS, 2010; TMA, 2020), however, there is a paucity of information on agricultural research data preservation in agricultural research institutions in Tanzania. This study, therefore, aims at closing this information gap.

Objectives of this study

The primary objective of this study was to investigate the roles of agriculture research data preservation methods/devices for enhancing data usage among researchers in Tanzania.

The specific objectives of this study

1. Examine the methods for the preservation of research data from research institutions in Tanzania.
2. Find out for how long agriculture researchers preserve their agriculture research data.
3. Determine factors influencing researchers in their choices of data preservation methods.

Significance of the study

The study will enlighten the agricultural research institutions to strengthen existing and adopt better methods for preserving the agricultural research data in short-term and long terms plans. Findings will create awareness in agriculture research institutions to see the need of strengthening the department of technology transfer to perform a coordination function to manage and preserve agricultural research data. The study shall also contribute to the existing body of knowledge by identifying data preservation methods, and factors that influence researchers on the choice of data preservation methods.

Literature review

This section organized literature related to the study topic. The literature review is organized around broad themes based on study objectives as follows: scientific data preservation methods, duration of data preservation, and factors that influence agriculture researchers in their choices of storage and preservation methods.

Research data preservation

Data can be in form of images, recordings, musical compositions, verbal communications, experimental readings, simulations, codes, textual, qualitative, and quantitative (Tripathi et al., 2017). While research data refers to the basic data which are generated by a human being through technological activities for example structure graphic data (SPSS, GIS), source code data, images data (JPEG, GIF, TIFF) (Adika and Kwanya, 2020; Zhao and Wang, 2015). In this case, agricultural research data can exist in the form of text documents, graphics, databases, spreadsheets, laboratory notes, films, photographs, test responses, and audio (Chigwada et al., 2017; Elsayed and Saleh, 2018). Furthermore, agricultural researchers generate several types of data such as crop breeding data, baseline survey data, genomic data, remote sensing data, and geographic information system (GIS) data (Dileepkumar, 2014; Kirub, 2016). Such important data generated are required to be preserved for future reuse by researchers and other stakeholders.

Data storage is among the activities in the data management life cycle. Adika and Kwanya (2020) define Research Data Management as a set of activities that include creation, organization, structuring, naming, baking up, storage, conservation, and sharing of data. Therefore, data storage is a process of recording or locating data in a medium that is managed by an individual or institution that allows access and sharing (Frederick et al., 2019). Such medium may include all storage devices, storage on a network, or systems, and storage in the cloud such as Amazon Web Services and Google Cloud Platform. The term data

preservation is defined as the permanent protection of research data that has been obtained from the finished project to guarantee data access for future use (Kruse and Thestrup, 2014) as cited in (Ng'eno and Mutula, 2018b). In the same vein, Mushi et al. (2020) pointed out that data preservation is one of the important activities carried out in the Research Data Management (RDM) life cycle though it has not yet been implemented in African countries including Tanzania.

Previous studies have shown that data can be preserved for the short term or long term. Literature shows that researchers prefer to store their data in short term during the project life cycle while long-term data storage was done after the end of the project (Dileepkumar, 2014; Tenopir et al., 2020). Researchers prefer to store their data temporally in their personal computers during the project life cycle. In line with this, Tenopir et al. (2020) exposed that researchers from Space and Planetary Science were satisfied to store data in their personal computers, departmental servers, or USB driver during the project life cycle. In the same study other respondents belonging to Marine Ocean (55.8%), Biology (50.9%), and Agriculture and Natural Resources (52.1%) were satisfied with long-term data preservation. Therefore, researchers are storing their data in short, and long-term data storage. The study also examined research data management in Kenya's agricultural research institutes (Ng'eno and Mutula, 2018b) findings revealed that the majority of respondents (20.2%) kept their research data for a period of more than 10 years, and those who kept their data between 5 and 10 years were 20.2%. Other respondents (23.4%) from this study did not know the length of time for research data preservation in agricultural research institutes.

Research data preservation methods

Previous literature has shown that researchers use different methods to preserve their data. Agricultural researchers from National Agriculture Research Systems (NARS) in Sub-Saharan Africa and lecturers at Strathmore University in Nairobi, Kenya were reported to use personal computers, flash, and external hard disks in managing their research data (Adika and Kwanya, 2020; Maru, 2004; Sielemann et al., 2020). Moreover, in a study that examine research data management in Kenya's Agricultural Research Institute (Ng'eno and Mutula, 2018a) it was found that (86.3%) preferred to preserve their research data on personal computers, and about 79% preferred to store their data into the hard drive, and 94.5% reported to preserve their data into their departmental servers, and institutional repository.

In line with this, in the study that examined data sharing, management, use, and reuse practices and perceptions of scientists worldwide (Tenopir et al., 2020) findings

exposed that 61.3% of respondents stored their data on personal computers, followed by 42.9% stored in institutional servers and 29.8% of respondents stored their data using their USB/external drivers. Likewise, in the web-based survey that investigated data preservation at the University of Kansas (Weller and Monroe-Gulick, 2014) findings showed that researchers preserved their data using electronic devices, and tools such as hard drives, CDs, and cloud-based storage like Dropbox or Google Drive and centralized server. This implies that the researcher had more options for data storage and preservation. However, these findings show that data stored on a flash drive and a personal computer was a bad practice while a good practice was storing data in institutional repositories and departmental servers. Data storage through personal devices was viewed as bad practice simply because it denies data access to others.

Literature has revealed that researchers prefer to preserve their data on a centralized data platform. The ICRISAT program which operates in Asia and Sub-Saharan Africa came up with innovative data preservation, publication, and web-based sharing platforms. Through this program, several data preservation platforms were initiated including Dataverse, Cloud-based M, AGROBESE, and Intergraded Breeding Platform (IBP) (Dileepkumar, 2014; Maru, 2004). These data platforms provided a chance for agricultural researchers elsewhere to access, share and reuse other researchers' data. In line with this, the European Biodiversity Network (EU BON) adopted a tool for data preservation, sharing, and description. In this network, several electronic data management platforms were established including Drupal Ecological Information Management System, Metacat and Morpho, Plazi Treatment Bank, Golden Gate Imagine, Plouto, and GBIF Intergraded Publishing Toolkits. These data preservation platforms enabled researchers to publish, share and access biodiversity data for use (Smirnova et al., 2016). These data platforms facilitated data sharing and easy data accessibility among members.

Moreover, research guideline from SUA and TARI headquarters stipulates that research data and other materials generated by researchers and students belong to the research institutions, and must be submitted to be preserved in the research institutions (DPRTC, 2019; URT, 2016). However, research guidelines do not tell something about how such important data are preserved for future usage. None of these previous studies managed to give a clear comprehensive understanding of research data preservation methods used in agricultural research institutions in Tanzania.

Choices for data preservation method

Several factors can influence researchers in the selection of data preservation methods.

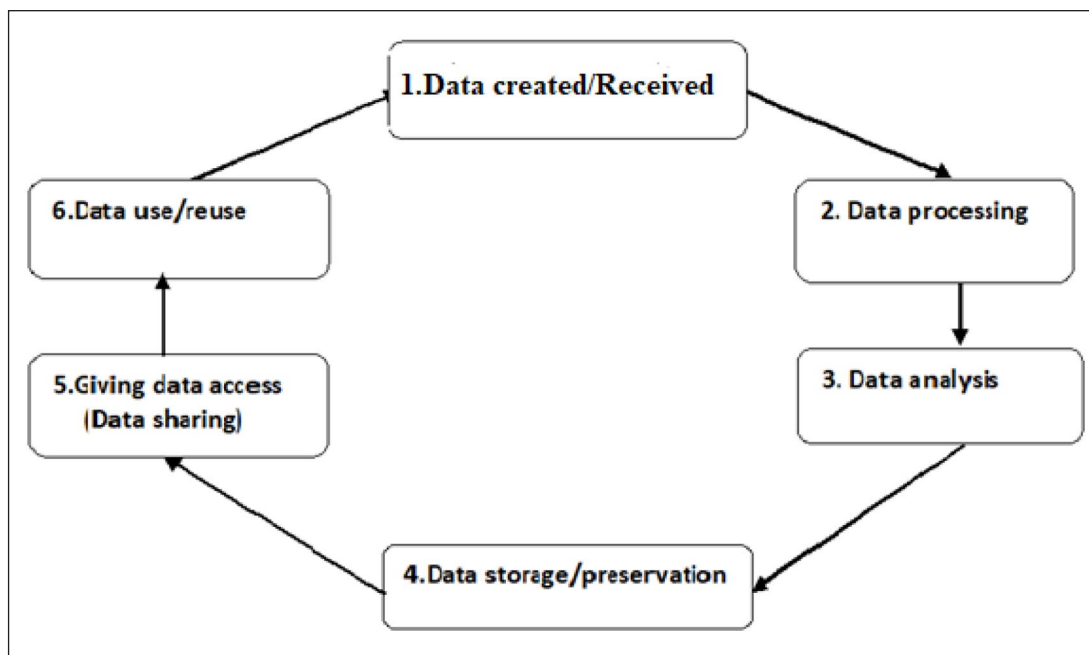


Figure 1. Agricultural research data life cycle model.
Source: Adapted and modified from Higgins (2008).

Previous literature has revealed that the availability of Information Communication Technology (ICT) tools and infrastructures in agricultural research institutions facilitated researchers to capture data, store, manipulate, analyze, preserve, and share their research data (Ng'eno and Mutula, 2018a; Zhao and Wang, 2015). Furthermore, technical support provided to researchers enhances researchers to choose the appropriate method for data preservation. Tenopir et al. (2020) revealed that the availability of data managers or data librarians who were readily available to assist researchers in their data preservations could facilitate researchers to preserve their data in long term by depositing their data into repositories beyond the project life cycle.

Furthermore, the cost of data preservation and management technology can encourage or discourage researchers to use a particular technology. If the cost of using the device is cheaper can influence researchers to use a certain dataset also if the cost is higher researchers will be discouraged to use the device or method to manage their data (Kirub, 2016; Musker and Schaap, 2018). Likewise, data preservation devices use electricity. The availability of reliable electricity is one of the factors influencing the choice of method to preserve data. Electricity is important in running electronic devices, and equipment. Unreliable electricity supply may hinder researchers not being able to use a particular data preservation equipment. According Barakabitze et al. (2015) reported that there were unreliable electric supplies to operate computers and other electronic storage devices in agricultural research institutions in Tanzania.

These previous studies have indicated the existence of some factors that influence researchers on the choice of method or device to use in preserving data. The methods varied depending on the duration of data preservation. Researchers preferred to preserve their data using personal preservation devices for the short term. While in some institutions researchers preferred to use formal or centralized systems to preserve their data for the long term beyond the project lifecycle. It is not known yet how agricultural researchers in Tanzania preserve their data during and beyond their project life cycle. Due to this scant documented evidence, there was a need to fill this knowledge gap.

Theoretical framework

This study adopted Data Curation Centre (DCC) Life cycle Model Figure 1 below. The Curation Lifecycle Model was proposed by the Digital Curation Centre (DCC) to promote a life cycle approach to the management of digital materials, to enable their successful preservation from the initial stage to their disposal or selection for reuse and long-term preservation (Higgins, 2008). The DCC model was useful to guide this study because the model explains that data are created, received, analyzed, preserved, shared, and finally data act as raw materials to generate new data Figure 1 below. The model is relevant to this study because it clarifies data preservation as one of the core activities in the data life cycle. Data preservation or management is a participatory activity it involve researchers, data managers, curators,

repository specialist, librarians, and project managers (Pouchard, 2016). The model expresses the value of data preservation thus without data preservation, data will not be available, easily accessed for used/reused in future research (Ng'eno and Mutula, 2018).

The Data Curation Centre (DCC) Lifecycle Model in Figure 1 below includes the data capture, processing, analysis, preservation, access, re-use, and transformation of research data (Ng'eno and Mutula, 2018a).

Data creation/capture

Data are created through the research process. Also, data may be captured by receiving data from other sources (Adika and Kwanya, 2020).

Data analysis

After data is created/received then data analysis is carried out. After analysis data can be preserved for future use and reuse while other data can be published in scientific works.

Data preservation

Data preservation means protecting data in a secured environment for the long term that will allow future access and reuse (Ng'eno and Mutula, 2018b). Researchers should choose better devices or methods for better data preservation. Better choice of devices or methods can allow data access and reuse in the future.

Data access

Data may be accessed from storage devices or methods. Data will be accessed when they are shared into accessible sources such as personal storage devices (computers), databases (data repositories), and libraries.

Data use/reuse

The preserved data may be used and reused if they were preserved on the devices or a method that allows data access and usage (Higgins, 2008).

Agricultural research data preservation conceptual framework

The conceptual framework in Figure 2 below indicates researchers create data during research activity or project (Tripathi et al., 2017). Also, researchers receives data from other sources to be used in the research process. Data obtained from research are analyzed, interpreted, and shared through publications while other data are preserved for future use/reuse. Researchers are required to have

better choices in the selection of data preservation devices or methods. The right choice of data storage and preservation devices/methods will ensure the data is effectively preserved and used. The data that has been preserved using a better storage device/method will enhance data access, sharing, use, and future reuse (Hawkins et al., 2018; Higgins, 2008).

Methodology

The study employed a descriptive and cross-sectional research design to investigate the role of research data storage and preservation for enhanced data usage among agricultural researchers in Tanzania. The survey was conducted from February 2020 to March 2022. The study used both qualitative and quantitative approaches in data collection and analysis. The qualitative approach was used to determine opinions and views on data storage and preservation. While the quantitative approach was used to examine the preservation methods and factors influencing researchers on the choice of storage and preservation methods/devices.

A combination of the two approaches helped a researcher to get relevant data for the study. The study was conducted in six agricultural research institutions in Tanzania. The ecological zones included in the study were randomly selected. These included; Eastern, Central, Northern, Southern, Southern Highlands, and Lake Zone. Tanzania Agriculture Research Institution (TARI) centers and SUA specifically the College of Agriculture and College of Veterinary Medicine and Biomedical Sciences were purposively selected to be included in the study. The selection of these research institutions considered the agricultural research institutions which had more than twenty (20) researchers at the institution. This number of researchers was enough to justify the data collection process.

The study included agricultural researchers. The total population of researchers in the study areas was 527. From this study population, the sample size of 227 was obtained using Yamane's (1967) formula for sample size calculation.

$$n = \frac{N}{1 + N(e)^2}$$

Where:

n= Sample size (227)

N= Total study population (527)

e= Level of significance (5%)

After obtaining the sample size of 227 then a representative sample per institution was calculated. This representative sample size per institution facilitated the determination of the total number of respondents to distribute the questionnaires. The representative sample per institution was

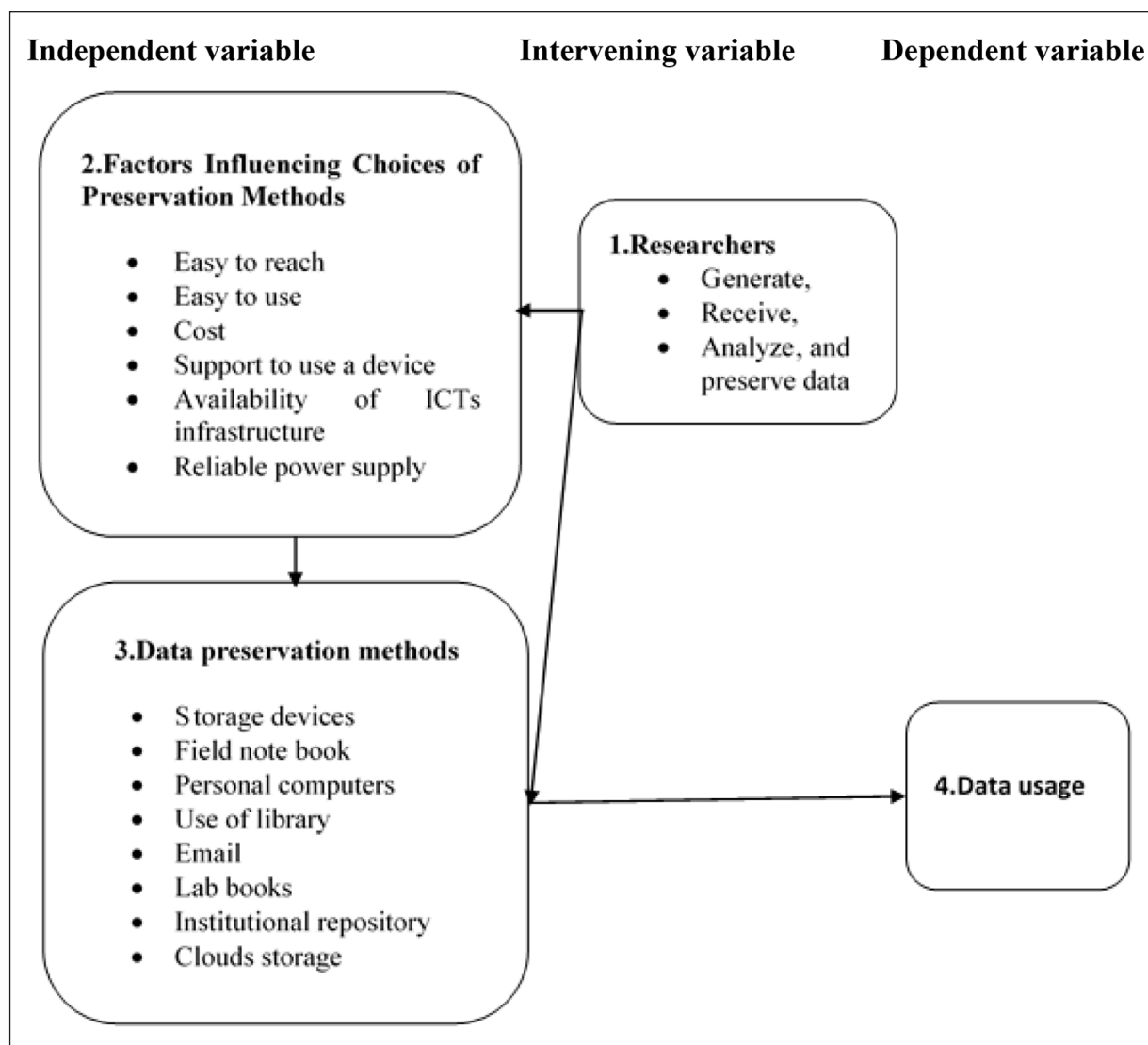


Figure 2. Agricultural research data preservation conceptual framework.
Source: Designed by researcher, 2022.

calculated (Table 1). From the sample size per institution, the respondents were randomly chosen to represent the 227 sample size. Eleven directors from TARI and one from SUA Post Graduate Studies were selected for key informant interviews. These were purposively selected because they were thought to be experienced and knowledgeable and would therefore provide in-depth insights about the topic.

Data collection involved the use of multiple data gathering techniques to investigate the methods use in data preservation among researchers. The cross-section survey method (questionnaire and interview) and focus group discussions were employed in the study. The questionnaire with both open and closed-ended questions was distributed to 227 agricultural researchers from SUA and 10 selected TARI centers. A total number of 204 questionnaires from

SUA and 10 selected TARI centers were correctly filled, returned, and used in this study (Table 1). The qualitative data were collected through interviews with key informants that lasted 15–20 minutes.

A total of two focus group discussions with agricultural researchers were conducted at Ilonga and Mikocheni centers, and each group had five participants. These research institutions were selected randomly from among several agricultural research institutions. The main purpose was to understand the types of research data preservation that are used in sharing agriculture research data. During the discussion respondents were free to express their views which were recorded by the researcher. In addition, secondary data was collected from different sources, including the SUA Research Guideline and Regulation of 2019 and the TARI Act of 2016.

Table 1. Sampling frame ($n=204$).

Zone	Institution	Total population	Sample size per institution (S) ^a	Returned questionnaire
Eastern	SUA	223	96	84
Lake	Ukiriguru	42	18	15
Southern highlands	Uyole	41	18	17
Southern	Naliendele	38	16	15
Northern	Selian	37	16	16
Eastern	Mikocheni	30	13	11
Eastern	Ilonga	28	12	12
Central	Makutopora	26	11	10
Eastern	Kibaha	22	9	9
Eastern	Katrin	20	9	8
Northern	Tengeru	20	9	8
	Total	527	227	204

Source: Field data, 2021, sample size per institution.

^aStudy sample size (SS) (227) times population per institution over total population (TT).

Pre-test

Before embarking on data collection, the instruments were pre-tested by 30 respondents in the College of Forestry, Wildlife, and Nature Conservation at SUA. The copies of the questionnaire were sent to them to check the overall structure of wording, presentation, and the relevance of the questions concerning the study objectives. Based on Cronbach's Alpha for testing the reliability of research tools, the test result gave an alpha of 0.943 which was above the 0.70 acceptable standards meaning that the tool was reliable to be used. A pre-test helps to identify errors that a researcher may have not foreseen during designing a collection tool (Plooy-Cilliers et al., 2014). Concerning ethical considerations, the researcher avoided any risk of harm during data collection and ensured the right of privacy and informed consent before engaging any respondent in the study.

Data collection and analysis

The data that was collected and analyzed included both qualitative and quantitative data. Qualitative analysis was accomplished using content analysis, the content analysis was used to analyze qualitative data that was obtained from key informant interviews, and Focus Group Discussions (FGDs). Quantitative analysis was accomplished by using statistical analysis, which is a technique that enables a researcher to examine quantitative data with the help of computer software packages for data analysis (Methew and Ross, 2010). The Statistical Product and Service Solution (SPSS) software version 22 and Excel software were used to carry out a descriptive analysis. The descriptive statistics results such as frequencies and percentages facilitated to organize and summarize data on methods used in data preservation, length of data preservation, and factors influencing researchers on choices of data preservation method.

Results and discussion

This part presents social-demographic characteristics of respondents such as sex, age, education level, research experience, and the research area of specialization. Also, this part presents findings based on study objectives; data preservation methods used to enhance sharing, and the factors influencing the researcher's choice of preservation methods.

Characteristics of respondents

The social-economic characteristics of the respondents are given in Table 2. Results show that there were twice as many males (67%) and (33%) female which shows that males dominated in agricultural research activities than females. Similarly, findings showed that 37% of respondents were in the age range of 31–40 years. Findings imply that there were more male agricultural researchers than female researchers within the institutions under study. This situation has been a tradition in many agricultural research institutes and in universities that have specialization in agriculture. It is not surprising to find more male employees than female employees. It is also worth noting that all of these institutions are dealing with scientific inventions and discoveries whose implementations require a background in science subjects. The findings are related to the study that found there were fewer women than men trained, recruited, and employed in the agricultural sciences and jobs related to science, technology, engineering, and mathematics (STEM) (Funk et al., 2018).

In terms of age distribution majority (37%) of respondents had an age distribution ranging from 31 to 40 followed by (32%) of respondents who had an age distribution ranging from 41 to 50 years. Respondents about (27%) with ages ranging from 51 to 61. This study found that the majority of the researchers in agricultural research institutions were aged 31 years and above. These results imply

Table 2. Demographic characteristics of respondents ($n=204$).

Gender	Frequency	Percentage
Male	137	67
Female	67	33
Age		
More than 61 years	2	1
51–60 years	55	27
41–50 years	66	32
31–40 years	75	37
21–30 years	6	3
Education level		
PhD	45	22
Master degree	123	60
Bachelor degree	36	18
Work experiences		
More than 8 years	163	79.9
6–8 years	17	8.3
3–5 years	20	9.8
Lesser than 3 years	4	2.0

Source: Field data, 2021.

that researchers from 31 years and above have been working on research projects for substantial years and therefore have generated many agricultural research data that are needed to be preserved for future use.

Findings in Table 2 indicate that majority (60%) of respondents had Master's degree while 22% of respondents had a PhD and 18% of respondents had a Bachelor's degree. In terms of education level of respondents, findings indicate that agricultural researchers are highly qualified for their work as the majority had a Master's degree and above in agricultural sciences. The respondents with this academic qualification have enough knowledge and skills in conducting research and preserving their research data.

Furthermore, findings in Table 2 below indicate that 87% of respondents were those with work experiences of more than eight (8) years. While about 11% of the respondents were those with research experiences between 3 and 5 years. About 8.3% of respondents were those with research experience between 6 and 8 years. And finally, 2% of respondents were those with lesser than 3 3 years of experience. Concerning work experience in research activities, findings signify that agricultural research activities require researchers who are well experienced to achieve the intended goal of improving agricultural production. Many agricultural researchers from SUA and TARI centers are well-experienced researchers they have a lot of research data that are preserved in their devices for future use and reuse. This demographic information was useful to indicate the suitability of the units for the study.

Findings in Table 3 below indicate the researchers' academic field of specialization. From the findings, the

Table 3. Researchers academic field of specialization ($n=204$).

Field of specialization	Frequency	Percentage
Botany, crop sciences, and horticulture	92	45
Animal science, aquaculture, and range management	16	7.8
Soil science and geological sciences	12	5.9
Veterinary medicine, surgery, and public health	11	5.4
Agriculture economics	11	5.4
Food technology, nutrition, and consumer sciences	10	4.9
Engineering sciences and technology	9	4.4
Agriculture biotechnology	9	4.4
Environmental science	8	3.9
Agro ecology	3	1.5
Public health and food safety	3	1.5
Natural resources management	2	1.0
Epistemology and laboratory management	1	0.5
Hydrology and water resources engineering	1	0.5
Total	204	100.0

Source: Field data, 2021.

majority (45%) of respondents indicated specializing in botany, crop sciences, and horticulture. In addition (7.8%) of respondents specialized in Animal Science, Aquaculture and Range Management, and Agricultural Extension and Community Development. Other respondents specialized in various fields as shown in Table 3 below. These results indicate that researchers from all research institutions specialized in one or more research fields. The primary role of research institutions has been stipulated in the Act of 2016 for TARI part 4(1) and from SUA Research Regulations and Guideline (DPRTC, 2019; URT, 2016).

Research data storage and preservation methods

The Data Curation Centre (DCC) Life cycle Model as shown in Figure 1 above emphasizes data storage and preservation as among the core activities in the DCC life cycle. Data preservation involves several data preservation methods. The finding in Table 4 below shows that more than 80% of respondents indicated they preserved their data using several methods/devices such as (flash sticks, external hard disks, CDs, and DVDs), use of field notebooks, and desktop computers found in departmental offices, and personal laptops. The findings in Table 4 below indicate that the prominent methods used by agricultural researchers to preserve their data include the use of storage devices such as (flash sticks, external hard disks, CDs, and DVDs), use of field notes books, and computer desktops and laptops. These findings imply that

Table 4. Data preservation methods existing and used in research institutions ($n = 204$).

Storage method	Frequency	Percentage
Storage device (flash sticks, external hard disk, CDs, and DVDs)	159	77.9
Field notes books	171	83.8
Computers desk/laptops	167	81.9
Institutional library	160	78.4
Storage device (flash sticks, external hard disk, CDs, and DVDs)	159	77.9
Use of email	107	52.5
Use lab books	91	44.6
Institutional database/repository	59	28.9
Use of clouds storage	50	24.5

Source: Field data, 2021.

researchers were preserving their data using informal ways that do not guarantee data availability and accessibility by all researchers. Findings have revealed that researchers from different research institutions preserve their data in their ways, and it happens the absence of the staff for any reasons may hinder data access among researchers. This practice is not health for the progress of science because researchers may lack important data for future research. However, some data preservation methods were not prominently used by agricultural researchers including lab books, institutional databases/repositories, and cloud storage. Among the reasons for the rare usage of these data preservation methods included low literacy in the use of data preservation methods such as cloud storage and the absence of data storage infrastructure such as institutional data repositories in the study area. SUA had an institutional repository that is not specified as a data repository however it hosts several types of research data in the form of research outputs including thesis/dissertation, and other electronic resources.

In addition through FGD from the Ilonga TARI center respondents reported:

We preserve our research data using our computer laptops/desktops, flash sticks, external hard disks, and CDs/DVDs. We submit our technical report to the Tanzania Agriculture Research Institute (TARI) General Director's Office for permanent preservation. Also, we normally submit research data to the project funder through their platform that is opened specifically for that project. At the end of the project sometimes you cannot be able to log in to that electronic platform. In our institute, we do not have a centralized database or data repository for research data preservation.

In addition through the interview from the TARI Ukiriguru center respondents reported:

I preserve my data in the departmental desktop computers as a common poor for permanent data storage. Also, at the end

of the research projects, I submit research data and other substantial research reports in hardcopies, and softcopies to the technology transfer department for permanent preservation.

In addition through an interview from SUA respondents reported:

Until now we do not have a data repository though we have an institutional repository used for preserving research outputs such as theses submitted to the institutions by researchers (academic staff), and students.

The use of personal laptops, other personal storage devices, and field notebook to preserve data is not a good practice this hinder data accessibility by other researchers. Moreover, it decentralizes data management to persons not trained as data storage managers (none data storage professionals). These findings are related to the previous studies which indicated that (22.6%), (8.8%), and (17%) of respondents for research institute A, B, and C respectively preferred to store their data in short term through their laptops, CDs, USB drives (Ng'eno and Mutula, 2018a; Tenopir et al., 2020). Furthermore, Aydinoglu et al. (2017) revealed that academics in Turkey used several ways to back up their data including CD/DVD/external hard disk and thumb drive, and cloud storage. There are some disadvantages of data storage on the hard disk and other magnetic devices. Data storage in magnetic tape and hard disk storage medium does not provide guarantees over time because it would become demagnetized and corrupted (Frederick et al., 2019). Therefore, the absence of research data repositories in agricultural research institutions in Tanzania might hinder researchers to have easy access, sharing, use, and reuse of their data.

Length of data preservation in research institutions

Table 5 below indicate the length of data preservation in agriculture research institution during, and beyond the project life cycle. The majority (74%) of respondents indicated they preserve their research data for more than 6 years. These findings imply that researchers preserved data for more than 6 years to enhance effective data use and future reuse. For example, researchers could prefer to use a computer, flash memory, and CDs to store data temporarily during the project life cycle while after the end of the project researchers could preserve their data in the office computers and also publish their data in funders database, online journals, submit research findings, and research materials to the institutions. Hawkins et al. (2018) revealed that cloud storage of agriculture data is a good way to store data for long-term stewardship.

Table 5. Length of research data preservation in research institutions ($n=204$).

Responses	Frequency	Percent
Lesser than 3 years	4	2.0
1–3 years	16	7.8
4–6 years	33	16.2
More than 6 years	151	74.0
Total	204	100

Source: Field data, 2021.

Factors influencing researcher's choice of preservation methods

Table 6 below show results from the descriptive data analysis showing the factors that influence the choice of data preservation method/device. Findings as shown in Table 6 below, indicated that the majority (50.3%) of respondents indicated agreed and (38.7%) indicated strongly agree that easy-to-reach storage method/device influenced agricultural researchers choice of method. These findings imply that researchers preferred to use storage devices, field notebooks, and personal computer laptops because they were easy to reach. It can be said that almost every researcher possesses a computer laptop and other storage devices so it is easy to reach and use such devices. These findings are related to a prior study (Weller and Monroe-Gulick, 2014) revealed that researchers preserved their data through hard drives, CDs, and university servers because they were easy to reach and use. Similarly, Tenopir et al. (2020) found that researchers were more satisfied with the availability of data in the repository because it was easy for them to locate and access data.

Furthermore, findings indicated the majority (52.5%) of respondents agreed, and (38.2%) of respondents strongly agreed that the ease to use of the device influenced researchers in their selection. The methods which are commonly used in data storage most of these are easy to use by researchers. The use of desktop computers in the offices, personal laptops, CDs/DVDs, flash memory, and external hard disk in most cases is easy to use by researchers. Furthermore, findings indicated that the majority (47.1%) of respondents agreed that the cost of using the device or methods influenced researchers on the selection of the device/methods to store their data. This implies that the cost of purchasing a data preservation system or device was observed to influence researchers in the selection and use of the method/device. The cost for the initial investment in some research data preservation devices/equipment or technologies is higher. Therefore, researchers' ability to purchase equipment or devices such as computer laptops, CDs/DVDs, flash memory, and external hard disks is cost-effective to researchers this influenced researchers to use such cost effective data preservation devices.

Equally, findings indicated majority (48.5%) of respondents agreed that adequate support provided on the use of data storage/preservation methods/devices influenced researchers to use such devices/methods. The support provided to the use of computer systems, use of flash memory, and other preservation methods influenced researchers in the decision to choose the method to use. For example, SUA has data managers who support researchers in all aspects of research and also the technology transfer office that coordinates research activities. Also, TARI centers have a technology transfer department responsible to coordinate research activities including data preservation and technology transfer. These findings are related to the previous study by Tenopir et al. (2020) who reported that the availability of data managers or data librarians assisted researchers in the procedures how to preserve their research data during, and beyond the project lifecycle.

Furthermore, findings indicated majority (33.8%) of respondents agreed that the availability of technical infrastructure in agriculture research institutions motivates researchers to use such a method/device. This implies that with the presence of institutional or departmental computers, personal laptops it was easy for them to use such available data storage and preservation devices. Findings are in line with prior study by Frederick et al. (2019) revealed that in the selection of data storage infrastructure one should consider the trustworthiness of the infrastructure in the data storage, select a repository that works well with FAIR principles, select a repository that allows granting different access condition to the stored files, and deposit your data to a repository that is accessible to users.

The aspect of reliable power supply in agricultural research institutions shows that (42.6%) of respondents agreed that reliable power supply can influence the use of a certain method/device in data preservation. The use of electronic data preservation devices such as computers, and CDs/DVDs requires a reliable electric supply to a particular research institution. Therefore, the availability of a reliable power supply influenced researchers to store and preserve their data in their desktop office computers, personal laptop computers, CDs, flash sticks, and other electronic devices. The Sokoine University of Agriculture, for example, has standby generators to ensure electricity is available all the time while it is different from some TARI centers where they do not have standby generators this might interfere with data preservation in case of power cut-off.

Recommendations

Based on the study findings the following are the recommendations.

Table 6. Factors that influence the choice of data preservation method/device ($n=204$).

Factors influence method/device selection	SA		A		U		D		SD	
	F	P	F	P	F	P	F	P	F	P
Device is easy to reach	79	38.7	103	50.3	14	6.9	8	3.9	—	—
Device/method is easy to use	78	38.2	107	52.5	12	5.9	7	3.4	—	—
Device/method is cost effective	51	25	96	47	36	17.6	19	9.3	2	1.0
Adequate support on use of device/method	44	21.6	99	48.5	27	13.2	28	13.7	6	2.9
Adequate data preservation infrastructure	27	13.2	69	33.8	44	21.6	58	28.4	6	2.9
Reliable power supply	33	16.2	87	42.6	25	12.3	52	25.5	7	3.4

Source: Field data, 2021.

SA: strongly agree; A: agree; U: undecided; D: disagree; SD: strongly disagree; F: frequency; P: percentage.

1. It is recommended that agricultural research institutions should establish research data preservation policies and regulations. The policies and regulations should state how agriculture research data will be preserved to ensure availability and accessibility. The policies and regulations for data preservation should be included in the institutional research guidelines.
2. It is recommended that the government should establish an agricultural research data bank to guarantee permanent availability and accessibility of data at all times when they are needed. The modern data preservation methods for example data repositories can be a central area that could enhance data sharing and usage among researchers. Also, the presence of centralized data management could ensure data is preserved permanently for a longer time after the end of the project.

Conclusions

Agriculture research data preservation is imperative in agriculture research institutions for the reason that if research data are preserved well, will promote data access, sharing, and usage. This study intended to investigate data storage, and preservation methods/devices to promote data access and reuse. Findings show that agriculture researchers store and preserve their research data during and beyond the project life cycle. The methods which are prominently used include the use of desktop computers, personal computer (laptops/desktops). This type of data preservation practice is not healthy for the progress of science simply because it does not guarantee data access, sharing, and future reuse. For example, in absence of concerned staff from the station for any reason would deny other researchers to have access and use of some important data. However, storage and preserving data using a personal device is a favorable practice for individual researchers as it permits easy access and use of data during and after the project. To foster the best agriculture research data storage, and preservation practices that could benefit researchers personally and the research institutions, there

is a need to adopt a formal way of preserving data. The presence of an agriculture research data bank can guarantee data availability, and accessibility at all times when they are needed. This study has contributed by proposing the data preservation conceptual framework that has been developed from the variables adopted from the Data Curation Center Model. This conceptual framework will help to enlighten researchers on the need to preserve their data for future use and reuse.

Acknowledgements

Sincere gratitude goes to my study supervisors from the Sokoine University of Agriculture and to my employer Jordan University College for giving me the opportunity to pursue further studies at Sokoine University of Agriculture (SUA).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Nolasko Victory Mwinami  <https://orcid.org/0000-0002-3375-7953>

References

- Adika FO and Kwanya T (2020) Research data management literacy amongst lecturers at Strathmore University Kenya. *Library Management* 41(6/7): 447–466.
- Adrian AM, Emison B, Musker R, et al. (2018) Open access and open data at push universities in partnership with global open data for agriculture (GODAN), (issue June 2018). PUSH University. Available at: <http://wp.auburn.edu/push/wp-content/uploads/2018/06/GODAN-report.pdf> (accessed 23 April 2022).
- Agrawal A (2017) Data roadmaps for sustainable development: Assessment and lessons learned. Available at: <https://www.data4sdgs.org/sites/default/files/servicesfiles/Roadmaps>

- AssessmentReport_Dec2017_final.pdf. (accessed 2 March 2022).
- Aydinoglu A U, Dogan G and Taskin Z (2017) Research data management in Turkey: perceptions and practices. *Library Hi Tech* 35(2): 271–289.
- Barakabitze AA, Kitindi EJ, Sanga C, et al. (2015) New technologies for disseminating and communicating agriculture knowledge and information: Challenges for agricultural research institutes in Tanzania. *The Electronic Journal of Information Systems in Developing Countries* 70(1): 1–22.
- Bhatia V, Stout S, Baldwin B, et al. (2016) Results data initiatives: Findings from Tanzania development gateway. Available at: <https://doi.org/developmentgateway.org/experience/results> (accessed 15 March 2022).
- Brouder S, Eagle A, Fukagawa NK, et al. (2019) Enabling Open-source Data Networks in Public Agricultural Research. CAST Commentary QTA 2019-1.p.20. <https://www.cast-science.org/wp-content/uploads/2019/05/QTA2019-1-Data-Sharing.p>
- Chidi C (2017) *Investigation of Factors Influencing Data Management in District Health Planning Process: A Case Study of Kibaha Town Council*. Dar-es-salaam: Muhimbili University of Health and Allied Sciences. Available at: <http://dSPACE.muhas.ac.tz:8080/xmlui/bitstream/handle/123456789/2160/Clara%20Chidi%20%20dissertation%20final.pdf?sequence=1&isAllowed=y> (accessed 21 March 2022).
- Chigwada J, Chiparausha B and Kasiroori J (2017) Research data management in research institutions in Zimbabwe. *Data Science Journal* 16: 31.
- Dileepkumar G (2014) Improving research data management and sharing: Experiences from ICRISAT. In: Proceedings of the winter school on livestock based livelihood options: current status, emerging issues and future scenario in combating agrarian crisis, New Delhi, India, 7 November 2014.
- DPRTC (2019) *Sokoine University of Agriculture Research Regulations and Guidelines, Sokoine University of Agriculture, 4th edn. Morogoro: Sokoine University of Agriculture*. Available at: <https://www.dprtc.sua.ac.tz/wp-content/uploads/SUA-Research-Regulations-and-Guidelines-2019.pdf> (accessed 12 October 2021).
- Eckes AH, Gubała T, Nowakowski P, et al. (2017) Introducing the brassica information portal: Towards integrating genotypic and phenotypic brassica crop data. *F1000 Research* 6: 465–516.
- Elsayed AM and Saleh EI (2018) Research data management and sharing among researchers in Arab universities: An exploratory study. *International Journal of Bioassays* 44(4): 281–299.
- FAO (2017) Country programming framework for united republic of Tanzania 2017–2020, prepared by the government of the united republic of Tanzania, ministry of agriculture, livestock and fisheries in collaboration with the food and agriculture organization of the united nations (FAO). Available at: <https://www.fao.org/3/bt133e/bt133e.pdf> (accessed 26 November 2021).
- Frederick C, Blumzon I and Pănescu AT (2019) *Data Storage*. London: Springer Available at: https://org/10.1007/164_2019_288 (accessed 25 May 2022).
- Funk BYC, Parker K and Calazza T (2018) Women and men in stem at odd over workplace equity: Perceived inequalities are especially common among women in science, technology, engineering and math jobs who work mostly with men. Pew Research Center. Available at: https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2018/01/PS_2018.01.09_STEM_final.pdf (accessed 20 May 2022).
- Granda P and Blasczyk E (2011) Data dissemination: Springer reference 688–713. Available at; https://org/10.1007/springerreference_64686 (accessed 5 May 2022).
- Hawkins E, Fulton J, Lee J, et al. (2018) Data fundamentals: Agricultural and biosystems engineering. Technical reports and white papers, p.25. Available at: https://lib.dr.iastate.edu/abe_eng_report/25 (accessed 28 August 2022).
- Higgins S (2008) The DCC curation lifecycle model. *International Journal of Digital Curation* 3(1): 134–140.
- Katabalwa AS, Bates J and Abbott P (2021) Potential opportunities and risks of sharing agricultural research data in Tanzania. *IASSIST Quarterly/International Association for Social Science Information Service and Technology* 45: 3–4.
- Kijazi A (2018) Data availability, exchange and gaps: A case study of Tanzania. Available at: https://library.wmo.int/doc_num.php?explnum_id=5380 (accessed 25 April 2022).
- Kirub A (2016) *Agricultural Research Data Management: Principles, Policy, and Practice*. Addis Ababa: Ethiopian Institute of Agriculture Research. Available at: <http://197.156.72.153:8080/xmlui/bitstream/handle/123456789/217/Data%20management-Book.pdf?sequence=1&isAllowed=y> (accessed 3 February 2022).
- Kruse F and Thestrup J B (2014) Research libraries' new role in research data management, current trends and visions in Denmark. *Liber Quarterly* 23(4): 310–335.
- Maru A (2004) ICT/ICM in agricultural research and development: Status in sub-Saharan Africa. Available at: <https://agris.fao.org/agris-search/search.do?recordID=XR2012210126> (accessed 3 January 2022).
- Mboera LEG (2015) *The Management of Health and Biomedical Data in Tanzania, Need for a National Scientific Data Policy, Directorate of Information Technology*. Dar es Salaam: National Institute for Medical Research. Available at: <https://slideplayer.com/slide/7625850/> (accessed 25 August 2022).
- Methew B and Ross L (2010) *Research Methods: A Practical Guide for the Social Sciences*. London: Longman. Available at: <https://www.pearson.com/uk/educators/higher-education-educators/program/Matthews-Research-Methods-A-Practical-Guide-for-the-Social-Sciences/PGM979381.html> (accessed 27 May 2022).
- Mulder N, Adebamowo CA, Adebamowo SN, et al. (2017) Genomic research data generation, analysis, and sharing – Challenges in the African setting. *Data Science Journal* 16: 1–15.
- Mushi GE, Pienaar H and van Deventer M (2020) Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania. *Data Science Journal* 19: 19.
- Musker R and Schaap B (2018) Global open data in agriculture and nutrition (Godan) initiative partner network analysis [version 1; peer review: 2 approved with reservations]. *F1000 Research* 7(5): 47.
- NBS (2010) *Dissemination and Pricing Policy: The NBS Recognizes That Timely Dissemination of Data to Specialized*

- Users for Research Purposes and Other Decision-Makers is Beneficial*. Dar es Salaam: National Bureau of Statistics. Available at: https://www.nbs.go.tz/nbs/takwimu/ACT/Dissemination_and_Pricing_policy220kb.pdf (accessed 17 March 2022).
- Ng'eno EJ and Mutula S (2018a) Research data management in Kenya's agricultural research institutes. Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (Information Studies) in the School of Social Sciences, College of Humanities, University of KwaZulu-Natal, Pietermaritzburg: South Africa. *Inkanyiso, Journal of Humanities and Social Sciences* 1: 333.
- Ng'eno EJ and Mutula S (2018b) Research data management (RDM) in agricultural research institutes: A literature review, submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (Information Studies) in the School of Social Sciences, College of Humanities, University of KwaZulu-Natal, Pietermaritzburg, South Africa. *Inkanyiso, Journal of Humanities and Social Sciences* 10(1): 28–50.
- Parliament of the United Republic of Tanzania (2016) Tanzania agricultural research institute (TARI) act no 10 of 2016, Dodoma: Tanzania. Available at: https://www.tanzania.go.tz/egov_uploads/documents/sheria_chapa_-Tanzania_Agricultural_Research_inst_sw.pdf (accessed 13 February 2022).
- Plooy-Cilliers A, Vavis C and Bezuidenhout R (2014) *Research Matters*. Cape Town: Juta and Company Ltd.
- Pouchard L (2016) Revisiting the data lifecycle with Big Data curation. *International Journal of Digital Curation* 10(2): 176–192.
- Sielemann K, Hafner A and Pucker B (2020) The reuse of public datasets in the life sciences: Potential risks and rewards. *Peer Journal* 8: e9954.
- Smirnova L, Mergen P, Groom Q J, et al. (2016) Data sharing tools adopted by the European Biodiversity Observation Network Project. *Research Ideas and Outcomes* 2: E9390.
- Tenopir C, Rice NM, Allard S, et al. (2020) Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS One* 15(3): e0229003–e0229026.
- TMA (2020) Tanzania meteorological authority (TMA) agrometeorological database. Available at: http://agromet.meteo.go.tz/storage/docs/TMA_Agromet_Database_Manual.pdf (accessed 3 February 2022).
- Tripathi M, Shukla A and Sonkar SK (2017) Research data management practices in university libraries: A study. *DESIDOC Journal of Library & Information Technology* 37(6): 417–424.
- URT (2016) Tanzania Agricultural Research Institute (TARI) act no 10 of 2016. Available at: https://www.tanzania.go.tz/egov_uploads/documents/sheria_chapa_-Tanzania_Agricultural_Research_inst_sw.pdf (accessed 13 February 2022).
- URT (2017) Information communication technology (ICT) policy for national bureau of statistics. Available at: https://www.nbs.go.tz/nbs/takwimu/ACT/ICT_policy_2017.pdf (accessed 22 May 2022).
- Weller T and Monroe-Gulick A (2014) Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech* 32(3): 467–482.
- Yamane T (1967) *Statistics: An Introductory Analysis*, 2nd edn. New York, NY: Harper, and Row.
- Zhao H and Wang J (2015) Development and current situation of agricultural scientific data sharing in China. In: *IFIP Advances in Information and Communication Technology*, vol. 452, pp.80–86. DOI: 10.1007/978-3-319-19620-6_10.

Author biographies

Mr. Nolasko Victory Mwinami is Assistant Lecturer at Jordan University College, Morogoro, Tanzania. Currently is pursuing PhD studies at Sokoine University of Agriculture(SUA), Department of Informatics & Information Technology (DIIT), College of Natural and Applied Sciences (CoNAS), Morogoro, Tanzania.

Prof. Frankwell W. Dulle is Senior Lecturer in the Department of Informatics & Information Technology (DIIT) at Sokoine University of Agriculture in Tanzania. He holds PhD in Information Science, MA (LIS) and BSc (Agric.). He is involved in teaching, research and consultancy activities at Sokoine University of Agriculture in Tanzania

Prof. Wulystan Pius Mtega is Senior Lecture in the Department of Informatics & Information Technology (DIIT) at Sokoine University of Agriculture in Tanzania. He holds PhD in Information Science from the University of South Africa. He is involved in teaching, research and consultancy activities. His areas of interest are knowledge management and Information Communication Technology (ICT) for development. He has written several publications in knowledge management and ICT for development.