RESEARCH PAPER

# Data Sharing Practices among Researchers at South African Universities

Siviwe Bangani and Mathew Moyo

North-West University, ZA

Corresponding author: Siviwe Bangani (Siviwe.Bangani@gmail.com)

Research data management practices have gained momentum the world over. This is due to increased demands by governments and other funding agencies to have research data archived and shared as widely as possible. This paper sought to establish the data sharing practices of researchers in South Africa. The study further sought to establish the level of collaboration among researchers in sharing research data at the university level. The outcomes of the survey will help the researchers to develop appropriate data literacy awareness programmes meant to stimulate growth in data sharing practices for the benefit of research, not only in South Africa, but the world at large. A survey research method was used to gather data from willing public universities in South Africa. A similar study was conducted in other countries such as the United Kingdom, France and Turkey but the Researchers believe that circumstances in the developed world may differ with the South African research environment, hence the current study. The major finding of this study was that most researchers preferred to use data produced by others but less keen on sharing their own data. This study is the first of its kind in South Africa which investigates data sharing practices of researchers from multi-disciplinary fields at the university level and will contribute immensely to the growing body of literature in the area of research data management.

**Keywords:** Data Management; Data Services; Data Preservation; Data Reuse; Legislation; Ethics

## 1. Introduction

Funding agencies are increasingly demanding that researchers archive their research data in open access repositories for long term preservation and sharing with others (Ross et al. 2018). Internationally, the top research funders such as the National Institutes of Health, National Science Foundation, European Commission, and the Organisation for Economic Co-operation and Development (OECD) encourage their grantees to develop data management and sharing plans (European Commission 2012; Buys and Shaw 2015). In response to these funding agency mandates, universities that are increasingly dependent on public and research agency funding have decided to formulate policies that require researchers to develop research data management and sharing plans for their research (University of Pretoria, 2017). As part of the global village, South Africa has not been immune from these developments. In 2015, the National Research Foundation (NRF) of South Africa, the primary research funder in the country, issued a statement that made it mandatory for researchers supported by the NRF to deposit their data to a reputable open-access repository (National Research Foundation 2015). According to the NRF, this will foster a culture of transparency and sharing of research data. Further, making research data readily available will foster inter-disciplinary research collaborations (Michener 2015), decrease research duplications, serve as a justification for research spending, ensure that the data can be used in ways that were not envisaged by the original data collector (Renaut et al. 2018), and protect the integrity of the research itself (Alter and Gonzalez, 2018). Universities and academic libraries in South Africa have taken note of these developments and are now moving towards open access research data management services and practices (Chiware & Becker 2018), which include the open sharing of research data. A study conducted by Onyancha (2016) showed that South Africa published close to 64% of all data published by Sub-Saharan African countries in the Data Citation Index hosted by Web-of-Science from 2009 to 2014. Despite this, there is currently scant evidence of the willingness, extent,

concerns and practices of sharing data by South African researchers (Patterton et al. 2018). The few studies conducted in South Africa in this area focus on data sharing practices of certain disciplines (Koopman & de Jager 2016). The current study has a broader aim. It seeks to establish the data sharing practices of researchers in South Africa from multi-disciplinary fields. This paper further seeks to establish the level of collaboration among researchers in sharing research data at university level in South Africa. In order to respond to this broader aim, this study has been divided into three objectives, which are to:

  ➢ Find-out the sources, formats, and size of research data produced by South African researchers.
  ➢ Determine the information assigned to research data produced by South African researchers.
  ➢ Establish the willingness, extent, concerns and practices of sharing data by South African researchers.

## 2. What is Research Data?

Scholars agree that a universal definition of data is difficult to craft. This is because the meaning of what is meant by data and the level of importance assigned to such data differs in different contexts and disciplines (Elsevier 2018). Borgman (2015: 29) defines data as entities used as evidence of phenomena for the purposes of research or scholarship. To Martone et al. (2018) data are the measurements, observations or facts taken or assembled for analysis as part of a study and upon which the results and conclusions of the study are based. In its definition, the OECD (2007) chose to focus on the types of data that are produced, which are: factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. In other words, data are a means to an end (Martone et al. 2018). They are means through which researchers validate research findings (Engineering and Physical Sciences Research Council 2018). The Engineering and Physical Sciences Research Council (EPSRC) (2018) adds that although the majority of scholarly data is in digital formats, research data in other formats need to also be considered. This study has adopted the definition of Martone et al. (2018) because of its all-encompassing nature as it covers all quantitative and qualitative research data types identified by the OECD (2007).

## 3. Data sharing

Martone et al. (2018) assert that data sharing is the publication of the primary data and any supporting materials required to interpret the data acquired as part of a research study. However, data sharing is not only about publishing the research data for easy access. Several studies have shown that sharing research data via emails and portable devices such as memory sticks or CDs is still practiced (Wiley 2014; Koopman & De Jager 2016; JISC, 2016). This caused Michener (2015) to simple define data sharing as the practice of making one's research data available or accessible for use by others. Tenopir et al. (2015) provides a broader definition of data sharing as the activity when a scientist or researcher makes their data available to others for use in research and other related activities. Dietrich et al. (2014) adds that researchers should not just share data among themselves without taking into account the legal and ethical considerations. Data sharing, therefore, is the practice of making research data available to others taking into account the legal and ethical implications associated with its sharing. According to the OECD (2007) and EPSRC (2018) data sharing has a number of benefits which include the following:

  · It increases the visibility of research data for citation and reuse by other researchers.
  · It reinforces scientific enquiry by ensuring that enthusiasts and sceptics alike are able to test, validate, and replicate research results.
  · It promotes new research and different ways of testing, and analysing research data.
  · It discourages unnecessary duplication of effort by researchers.
  · It saves financial and other resources that are wasted when similar data sets are created by different researchers.
  · It helps enable new discoveries from old research data sets.

## 4. Regulatory framework for sharing data in South Africa

Ethics and legislation are a major concern for data sharing advocates and sceptics alike (Sieber 2005; Dietrich et al. 2014). As such, researchers have an obligation to protect the welfare, rights and confidentiality of participants (Sieber, 2005; Seto and Luo 2007). Due to such ethical concerns, sharing certain types of data may be restricted, prohibited or harmful to the researchers, institutions, individuals or the society at large

(Research Data Alliance and The Committee on Data for Science and Technology, 2016). The Research Data Alliance and The Committee on Data for Science and Technology (2016) reminds researchers that they have an obligation to understand the applicable laws in their jurisdiction when sharing data. There are instances where ownership of the data itself is not very clear. In the case of academia, Cleary et al. (2013) point out that it is not always clear as to who owns the raw or primary data of a funded study that was abandoned by a postgraduate student in pursuit of an academic qualification. The same can also be said about personal medical data of patients in a private practice. It is not clear whether the doctor, practice, patient or even the state owns the data provided by the patients. In some instances sharing data may lead to more harm than benefits for the researchers, institutions, participants, and the society. Sharing trade secrets, for example, may cause harm to the creators and companies affected (Research Data Alliance and The Committee on Data for Science and Technology, 2016). In sharing data, researchers should take applicable legislation and regulation into account (Sieber 2005, Seto and Luo 2007, Research Data Alliance and The Committee on Data for Science and Technology, 2016). The Research Data Alliance and The Committee on Data for Science and Technology (2016) identifies five principles to ensure that research data is legally interoperable. Those are to: facilitate the lawful access to and reuse of research data, determine the rights to and responsibilities for the data legislation, balance the legal interests, state the rights transparently, promote the harmonization of rights in research data, and provide proper attribution and credit for research data. Legislation plays a paradoxical role when it comes to data sharing. It can be both an enabler and a barrier. In order to get a better context within which South African researchers (do not) share data, this section discusses the legislative and policy instruments that enable and restrict data sharing in the country. It is very important to note that the South African legislation does not have an explicit data legal framework like data protection laws and other legislation instead there is legislation that contains some limited data protection provisions (Roos, 2008).

### 4.1. The Constitution of the Republic of South Africa, 1996
The Constitution of the Republic of South Africa "the Constitution" is the supreme law of the country (South Africa 1996). Chapters 32 and 14 of the Constitution that are of particular relevance to information sharing. Chapter 32 provides South Africans, including researchers, a right to access information held by the state or another person especially if the information is used to exercise or protect a right. The promotion of Access to Information Act, 2000 (PAIA) is such a legislation. Like all rights in the Constitution of the Republic of South Africa, the right to information is not absolute. It has to be balanced with other rights in the Constitution such as the right to privacy in Chapter 14 (Howie et al. 2006). According to Howie et al. (2006), privacy is a valuable aspect of humanity. This is why researchers must take good care to protect it. Chapter 14 of the Constitution provides citizens of South Africa with the right to privacy which includes the right not to have their communications infringed upon.

### 4.2. The promotion of Access to Information Act, 2000 (PAIA)
PAIA was enacted to fulfil the obligation to give effect to the right to information regulated in Chapter 32 of the Constitution of the Republic of South Africa. Though PAIA does not explicitly refer to data, it is used to access information held by the state or persons in South Africa (South Africa 2000). Crucially, it also regulates matters connected to access to information, and seeks to ensure transparency and accountability. Internationally, PAIA is in line with Guideline 1D which requests that data (information) should be made available equitably to all users (Research Data Alliance and The Committee on Data for Science and Technology, 2016). It is the view of the current researchers that PAIA as the biggest enabler of information access and sharing also extends to the sharing of data associated with the information as also pointed out by Roos (2008).

### 4.3. The Protection of Personal Information Act, No 4 of 2013 (POPI)
Enacted to fulfil the right to privacy in Section 14 of the Constitution of the Republic of South Africa, 1996, the POPI Act serves as the biggest barrier to sharing data especially if the data will infringe on the privacy of people (South Africa 2013). It sets conditions under which data can or cannot be shared. This Act provides that research subjects are to be notified when personal information is collected about them. Under this Act research subjects have a right to object to the sharing of their data.

### 4.4. Electronic Communications and Transactions Act, 2002 (ECTA)
Among the purposes of ECTA, is to prevent abuse of electronic information and data (South Africa, 2002). Chapter 50 and 51 of the ECTA protect the privacy of parties involved in electronic communication, storage and sharing. Chapter 51 requires a data controller or holder to seek permission before collecting, collating, processing or disclosing data to third parties. The data may only be collected, stored, or shared electronically

for legal purposes. This places an obligation to researchers and data users to ensure that the data they share is requested and used for legal purposes. This is in line with Guideline 2B of the Research Data Alliance and The Committee on Data for Science and Technology (2016) which also places this responsibility to research data users who must abide by the rights applicable to the collection of the data. After the reasons for requesting electronic data is furnished, the receiver is not allowed to use the data for any other purpose/s than the one disclosed to the data controller or holder. The Act further advises against the use of electronic data that discloses personal information.

### 4.5. National Health Act, 2004 (NHA)
Chapter 9 of the NHA is dedicated to the conduct of national health research and information. Section 71 outlines the procedure to be followed when experimenting with human subjects while Sections 72 and 73 deal with the ethics of health research including data sharing. Crucially, Section 73 calls for every institution, health agency and health establishment where research is conducted to either establish a health research ethics committee or have access to a nationally registered health research committee. Most universities in South Africa are obliged to form these ethics committees in order for them to conduct health research. These committees also decide on the ethics of sharing data with other researchers.

### 4.6. The relevant South African intellectual property laws
Fair trade is an important ethical concern in data sharing (Parker, 2015). South African researchers should also take into account several intellectual property laws that act as an impediments to sharing data including the Trade Marks Act, the South African Copyright Act, the Patents Act, the Designs Act, and the Plant Breeders' Rights Act. The South African Copyright Act (South Africa, 1978) and Trade Marks Act (South Africa, 2008) protects writers and creators of published works from having their works distributed without their consent. In order to share trademarked and copyrighted data, researchers may have to contact the holders of such rights who may refuse or approve their requests. In respect of Section 65 of the Patents Act, researchers should take good care not to infringe the Act as this may lead to stiff penalties. Though not explicitly stated in the Act, sharing data that comes or could lead to the infringement of a patented product may be in conflict with the provisions of the Act. The Plant Breeder's Rights Act (South Africa, 1976) provides for the registration of new varieties of plants by the plant breeders as well as protects their rights to breed such plants. Researchers dealing with data involving plants should consider whether the data may lead to infringement of plant breeder's rights before sharing the data.

### 4.7. The NRF Open Access Statement as an enabler of data sharing
The NRF is a statutory body entrusted by the South African government with developing, supporting, and funding research in the country (South Africa 1998). In January 2015, the NRF issued a statement in support of open access scholarship. This statement, which became to be known as the NRF Open Access Statement came into effect in March 2015. It is generally seen as the biggest enabler of open science in the country (Koopman and De Jager 2016). Essentially, the statement mandates that research published using public funds must be made available in accredited open access repository. With regards to data, the statement calls for the beneficiaries to also share the research data emanating from the funded research to an accredited open access repository with the provision of a Digital Object Identifier for future citation and referencing. In their research data management policies, universities have been quick to respond to the NRF Open Access Statement. The University of Pretoria (2017), for example, explicitly refer to this statement in their research data management policy.

### 4.8. Research ethics committees of universities
In order to deal with the issues associated with ethics in research, universities in South Africa formulate policies that guide the conduct of research and data (University of Johannesburg 2007; North-West University 2018). The common theme among the policies is the respect for research subjects, deciding on the cost-benefit-analysis of conducting research, crediting sources used in research, and transparency. Further, the ethics committees of universities oblige all researchers to consult and obtain ethical clearance before their research proposals are approved by the universities.

## 5. Literature review
The literature review of this study covers five themes in line with the objectives. It focusses on previous studies that dealt with sources of research data, formats of research data, size of data produced by researchers, information assigned to research data, and the willingness, extent, concerns and practices of sharing

data among researchers. This is to assist the researchers to relate, compare and test the present study to the relevant research conducted in the field. In order to make the sharing of data seamless, the data should be stored in such a way that it is Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al. 2016). The first three objectives of this study sought to discover if the South African researchers follow the FAIR principles in storing their data.

According to Wouters and Haak (2017) for the research data to be shared it must first be managed, stored, and curated. It is in that context that it becomes important to understand the origins of research data in order to answer the question of where the researchers get their data. A Master's thesis by Kvale (2012) at the Norwegian University of Life Sciences found that about half of researchers make use of data produced by researchers in their own field while about 33% of researchers also use data from researchers in other fields. A study of the research sharing practices of researchers commissioned by Elsevier found that 43% of researchers rely on data shared to them from outside their teams (Wouters and Haak 2017). This shows that the practice of using data produced by others is common among researchers but more so among those that are in the same field. Research data can be stored in many formats including standard office documents, images, PDF files and other formats (Kvale, 2012). Kvale (2012) found that at the Norwegian University of Life Sciences, standard office documents were the most common data formats produced by 82% of researchers followed by images (55%), PDFs (49%), device specific raw data (42%), scientific and statistical formats (35%), and plain text (33%). Another study by Buys and Shaw (2015) found that spreadsheets made 68% of the data followed by structured data (58%), text (74%), and images (52%). A recent study by Elsayed and Saleh (2018) also took into account the types of data produced by researchers. Elsayed and Saleh (2018) found that the kind of documents produced were text (85.5%), spreadsheets (65.5%), images (56.4%), and statistical data (48.7%).

Data scientists are also interested to know the size of research data produced by researchers. This is because the size of data affects its management, publication and sharing (Chen and Wu 2017). Chen and Wu (2017) determined that 57.15% of chemistry researchers produced data in the gigabytes, followed by 29.41% who indicated that they produced data in the megabytes, 11.76% produced terabytes, while only 1.68% produced petabytes. In Buys and Shaw's (2015) study, 9% of researchers expressed that they required space for less than 1 Gigabyte, 42% required space in the Gigabytes, 15% in the terabytes, while 0% indicated a requirement for space in the petabytes. In Kvale's (2012) thesis it was reported that half of researchers do not assign any additional information to their data, 35% assigned administrative information while 30% assign technical information, while 16% assigned both the administrative and technical information. Literature shows that researchers have different attitudes, concerns, and practices when it comes to data sharing. Tenopir and others (2011) researched the attitude of researchers towards sharing data. Seventy-eight percent of researchers either strongly or mildly indicated willingness to share data. Tenopir and others (2011) determined that 74.9% of researchers admitted to sharing their data with others. However, 75% of researchers also expressed concerns about sharing their data because it may be misinterpreted. Another study on data sharing practices by Buys and Shaw (2015) determined that 70% of researchers were unwilling to share their data. In Buys and Shaw (2015), privacy or protection were the major concern for 37% of the researchers while intellectual property rights were a concern for 21% of researchers and 19% of researchers had concerns that other researchers would not be interested to their research. A study of the Wellcome Trust researchers by Van den Eynden et al. (2016) determined that 51% of researchers share their data through repositories with established researchers sharing significantly more datasets than emerging researchers. There were no significant differences with regards to data sharing between researchers based in the UK and those that are in other countries. The researchers expressed that their major concern was that data may be misused or misinterpreted as a major impediment for them to share data with others. Wouters and Haak (2017) also studied the willingness of researchers to share data and found that although 73% of them agree that they will benefit from other researchers sharing their data, only 64% of them were willing to share data and 65% have actually shared data. Another study by Elsayed and Saleh (2018) found that 64.4% of researchers shared their data with other researchers. Concerns with data sharing identified in this study were data privacy and confidentiality (57.1%), the fact that it takes time and effort (37.3%), intellectual property rights (31.3%), and technical issues (29%). Researchers have identified barriers to data sharing that are peculiar to South Africa. In a study reporting on a workshop about sharing data to improve health in Africa, Parker (2015), a UK-based researcher, point out that South African researchers identified some impediments to data sharing in South Africa. Those are lack of incentives to share data, lack of a culture or precedence of data sharing in the country, lack of infrastructure for data management and curation and insufficient allocation of funds to research projects for data management and curation. These impediments may also be applicable to other third world countries.

## 6. Methodology

This survey formed part of international studies of a similar nature. The researchers used a pre-prepared online questionnaire to determine data management and sharing practices in South Africa. The online questionnaires were distributed via emails to academics, researchers, Masters and PhD students at the following universities in South Africa: Venda, North-West, Sol Plaatjie, Cape Peninsula University of Technology, and Tshwane University of Technology in June and July 2018. One hundred and forty-one responses were received of which 129 were found to be useable. Twelve questionnaires could not be used as the information supplied was incomplete. Admittedly, the response rate was lower than anticipated. However, the researchers decided to use the data as the response rate was in line with several studies of a similar nature elsewhere. Buys and Shaw (2015), Wouters and Haak (2017), and Elsayed and Saleh (2018) also had lower response rates of 6.4%, 2.3% and 8% in studies similar to this one. Schöpfel and Prost (2016) point out that some of the similar studies' response rates are partially unknown as no one knows the whole population of the studies. Bezuidenhout and Chakauya (2018) only managed 100 responses from 13 countries. Kuipers and Van der Hoeven (2009) could only manage 46 respondents from the Netherlands, 50 from France, 51 from Italy, 67 from Germany and 129 from the United Kingdom. In this study, twenty-five questions were asked and the respondents were given an opportunity to select from a number of options. The questions contained single-response and multiple-response multiple-choice questions. All responses were sent to the original creators of the survey. This study focused on the responses to 10 questions that related to the data sharing practices of the researchers including demographic information such as primary role, age and discipline of the researchers. Ethical clearance to conduct this study was requested and approved by the North-West University Ethics Committee. The NWU ethics number assigned to this study was IRB-2017-11-002. A letter of request with a link to the online survey was e-mailed to all the public universities through the relevant library directors requesting them to distribute the online questionnaire to researchers and postgraduate students in their universities. Cooperating library directors responded indicating the processes that the researchers needed to follow as per their ethical policies and the researchers complied with such policies where applicable. Other directors simply proceeded to distribute without asking for further compliance particularly given that the NWU ethical clearance was attached. The 10 questions used in this study were converted into MS-Word and attached as Appendix A.

### 6.1. Data analysis

The researchers did not receive the responses directly as they were sent to the original creators of the questionnaire. As part of the prior agreement, the researchers requested the questionnaire responses from the creators who are based in Turkey and France. The data was received in Excel spreadsheets which allowed for easy analysis. After data cleaning, the responses were then sorted accordingly and graphs created. The findings of this study are presented and reported in the form of graphs, percentages, and aggregates.

### 6.2. Respondents' profile

The first three questions of the survey were about the profile of the researchers. The researchers were asked to indicate among others their current primary role, age, and discipline. Ninety-eight respondents indicated that they were masters or PhD students, 28 were researchers and academics, while 3 did not indicate their primary role. The domination of postgraduate students in this survey is in line with similar surveys in other countries (JISC, 2016). In terms of age 51 were between the ages of 18 to 25. Fifty-two were between the ages of 26 to 35. Twelve were between 36 to 45. Nine were between the ages of 46 to 55, and 5 were 56 to 65. This is in line with other surveys like Buys and Shaw (2015) who also noted similar trends where the majority of respondents were postgraduate students. Respondents were given 37 choices to choose their field of study. However, these were condensed into seven faculties in this study for easy analysis of data. As such, there were 59 Natural and Agricultural Sciences respondents, 27 were Humanities, 19 were Economic and Management Sciences, 9 were Education, 7 belonged to Law, 4 belonged to the Health Sciences, 1 was an Engineering researcher, and 3 did not indicate their discipline.

## 7. Findings and discussions

In order to understand the sharing practices of researchers it is important to first establish as to who creates the data they are using. The first question of the study was to find out the sources of research data for South African researchers. There were four choices given to the researchers and the results are reflected by **Figure 1** below.
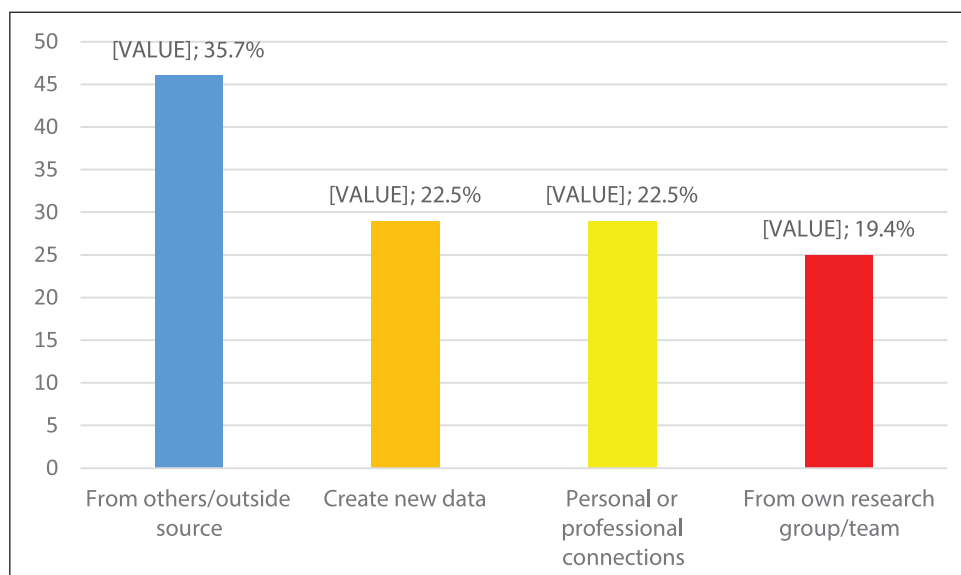
**Figure 1:** Sources of research data for RSA researchers.

When asked as to how they normally source their data most researchers indicated that they use research data created by others. Only close to 23% of researchers create new data. More than 77% of the researchers got their data from elsewhere either through personal or professional connections, from own research group or team, and from outside sources. This places a greater responsibility to researchers to ensure the highest integrity, easy access and availability of their new data so that it can be reused by other researchers. These results are comparable to Kvale (2012), Van den Eynden et al. (2016) and Renaut et al. (2018). Kvale (2012) in particular found that the majority (83%) of researchers use data from others rather than create new data at some stage in their lives even though the extent of reuse of the data differ according to discipline. Van den Eynden et al. (2016) found that 49% of researchers sponsored by the Wellcome Trust obtained data from colleagues or collaborators.

The second objective of this study was to identify the formats of research data produced by South African researchers. The researchers were given 13 options to select. **Figure 2** reflects the formats of the research data produced by South African researchers.

Sixty-four percent of researchers indicated that they produce data in standard office documents such as text, spreadsheets, presentations and others. Close to 36% claim to produce structured scientific and statistical data formats such as SPSS, GIS, and FITS. Close to 19% produce Internet and web-based data such as webpages, e-mails, blogs, and social network data. The percentage for images such as JPEG, GIF, TIFF, PNG etc. stood at close to 15%. Archived data (such as ZIP, RAR, JAR, etc.) was produced by only one (0.8%) researcher. Previous studies (Kvale 2012; Buys and Shaw 2015; Elsayed & Saleh 2018; and Bezuidenhout & Chakauya 2018) also found similar to the current one. All these studies found that the majority of researchers produced standard office documents. Kvale (2012), for example, determined that 83% of data produced by researchers at the Norwegian University of Life Sciences were standard office documents while Buys and Shaw (2015) found that 66% of data produced were text. Elsayed and Saleh (2018) also found that 85.5% of research data produced at Arab universities were text documents. Kuipers and Van der Hoeven (2009) on the other hand showed that 25% of European researchers were producing archived data (ZIP, RAR, JAR, etc.) by 2009. This does not compare well with this study which shows the percentage of archived data to be less than 1.

The amount of data produced by researchers is very important for purposes of determining their sharing practices (Chen and Wu 2017). The graph in **Figure 3** reflects the size of documents produced by South African researchers.

A plurality of researchers (37%) produce data in the megabytes while 24% produce gigabytes, and only 9% gigabytes. Twenty-four percent of researchers are not sure about the size of data they produce. The results in this study are slightly different to Buys and Shaw (2015) and Chen and Wu (2017) where the majority of researchers expressed that they produced data in gigabytes. These results may mean that the fears of huge server space to store research data may be unfounded. The majority of data produced by researchers
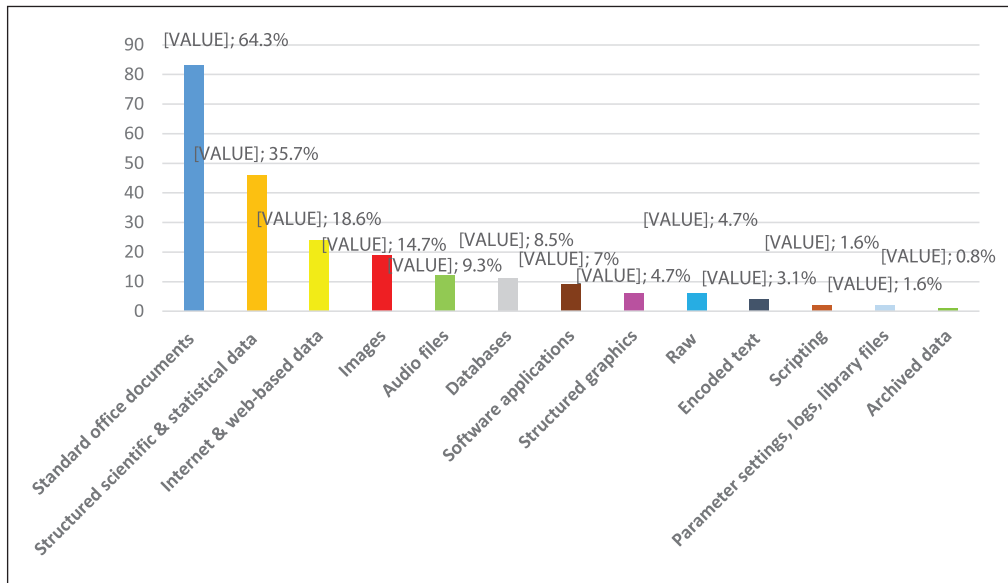
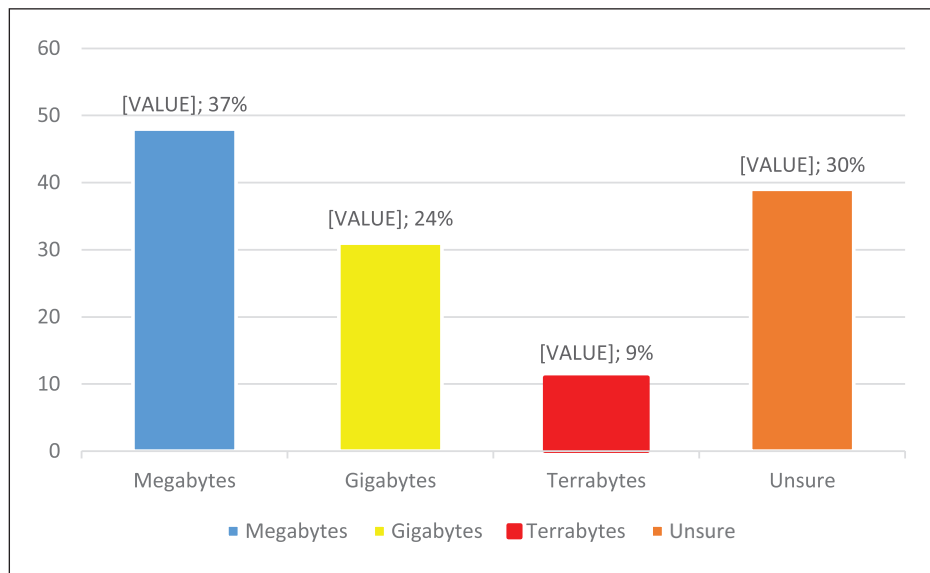**Figure 2:** Formats of research data produced by RSA researchers.



**Figure 3:** The amount of data produced by South African researchers.

do not require a lot of space. The high percentage of researchers who were unsure about the amount of data they use was significantly different from that of Kuipers and Van der Hoeven (2009) who found that 17% of European researchers were not sure of the amount of data they produce. This could be attributed to the fact that the majority of researchers in this study are at an early stage of their research careers.

To further determine the possibility for sharing and reuse of the data, the fourth objective sought to determine the information South African researchers assign to research data. For this question, the researchers were given five options. **Figure 4** determine the information assigned to research data produced by South African researchers.

More than 30% of researchers conceded that they do no label their research data files accordingly. Close to 70% of researchers label their data files in one way or the other with very few (close to 8%) managing to label them perfectly. Just more than 33% supply administrative information (e.g. creator, date of creation, file name, access terms/restrictions) while just more than 19% provide the technical information such as file format, file size, software/hardware needed to use the data. Close to 18% of the respondents indicated that they provide the discovery information such as creator, funding body, project title, project ID, and keywords in their files. These results are different from those of Kvale (2012). The majority of researchers in this
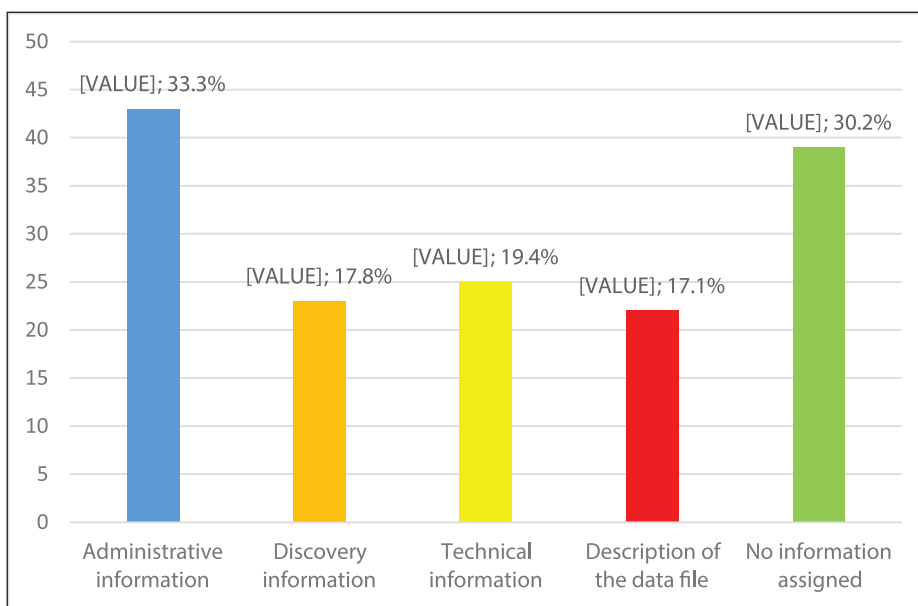
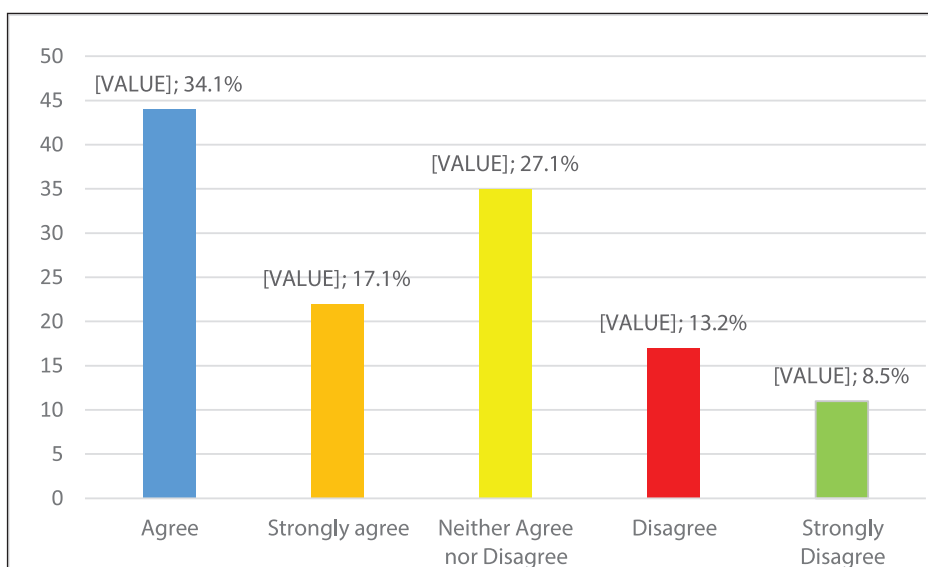**Figure 4:** Information assigned to research data.



**Figure 5:** Willingness of researchers to share their data.

study label their research data files in one way or the other unlike in Kuipers and Van der Hoeven (2009) and Kvale's (2012) study where 70% and 50% of the respondents did not assign any additional label to their files. However, it must be pointed out that things may have changed in Europe since the above studies as lots of advocacy and training around research data management happened since 2009 and 2012. In the case of South Africa, this points to some level of success of earlier interventions by librarians to assist researchers with their data management practices at South African universities (Chiware and Becker 2018).

Figure 5 reflects the willingness of researchers to share their data. When asked if they will be willing to share their data with other researchers just more than 51% of researchers agree or strongly agree with this. This is despite the fact that more than 76% of them use data from other sources. More than 27% choose to remain non-committal meaning that they may be open to persuasion. Close to 22% would not want to share their data at all with close to 9% stating that they would strongly disagree with sharing their data. These results differ to those of Tenopir and others (2011). Tenopir and others (2011) determined that the vast majority of researchers were willing to share their research data. More than 80% of researchers in Tenopir and others' (2011) study either agree or expressed a strong agreement that they would be willing to share data. The results of this study, however, compare well with those of the JISC (2016) study and Patterton et al.

(2018). The JISC (2016) study determined that 10% of researchers were not willing to share their data. In the JISC study, only 33% of researchers were already sharing data while 35% expressed a willingness to share data in future. Patterton et al. (2018) on the other hand determined that 12% of emerging researchers were not willing to share their data.

Researchers also expressed concerns about sharing research data as shown in **Figure 6**.

Just more than 30% of researchers expressed no concerns about sharing data while just more than 22% expressed legal and ethical concerns followed by those concerned about possible misinterpretation of their data (14.7%), misuse of the data (11.6%), lack of appropriate research data management policies (8.5%), fear of losing the scientific edge (close to 7%) and lack of resources to share the data (5.4%). In the current study the legal and ethical issues were the biggest concern while fears of misinterpretation or de-contextualisation of the research data, and its outright misuse were the second and third biggest concerns. In Kvale's (2012) study, the misuse of data was the biggest concern at 48% followed by the ethical and legal issues at 41%. The fear of losing the scientific edge stood at 30% while in this study it is close to 7%. This may point to the uniqueness of the researcher concerns about research data sharing by country. It is possible that researchers in each country have unique concerns about sharing their data. However, Kvale's (2012) study was at a single institution in Norway which may not represent all universities in that country. In the case of Europe, legal issues (41%) and misuse of data (41%) were the major concerns of researchers by 2009 (Kuipers and Van der Hoeven, 2009). However, this may have changed for better or worse since the adoption of the General Data Protection Regulation by the European Union in 2016.

This study was also interested to find out the current data sharing practices of South African universities as reflected in **Figure 7**.

When asked about their current research data sharing practices, close to 19.4% of researchers in South Africa indicated that they share their data with others. Confusingly, close to 26% admitted to sharing their data with researchers in the same team which may mean that some researchers may have misunderstood the first question. Only close to 14% of researchers admitted to sharing their research outside their institutions. This may point to limited inter-institutional collaborations in South Africa as also found by Maluleka and Onyancha (2016). Additionally, this may point to a lack of trust among researchers of those who are not part of their inner circle. These results differ with those of Tenopir and others' (2011) who determined that 74.9% of researchers admitted to sharing their data with others. This points to different data sharing cultures between South Africa and possible other countries and the United States of America where Tenopir et al. (2011) conducted their study. This study confirms the findings of Elsayed and Saleh (2018) and other researchers that of a gap between researchers' sharing practices and their own usage of data. While researchers are happy to make use of data produced by others, they are averse to sharing their own data.
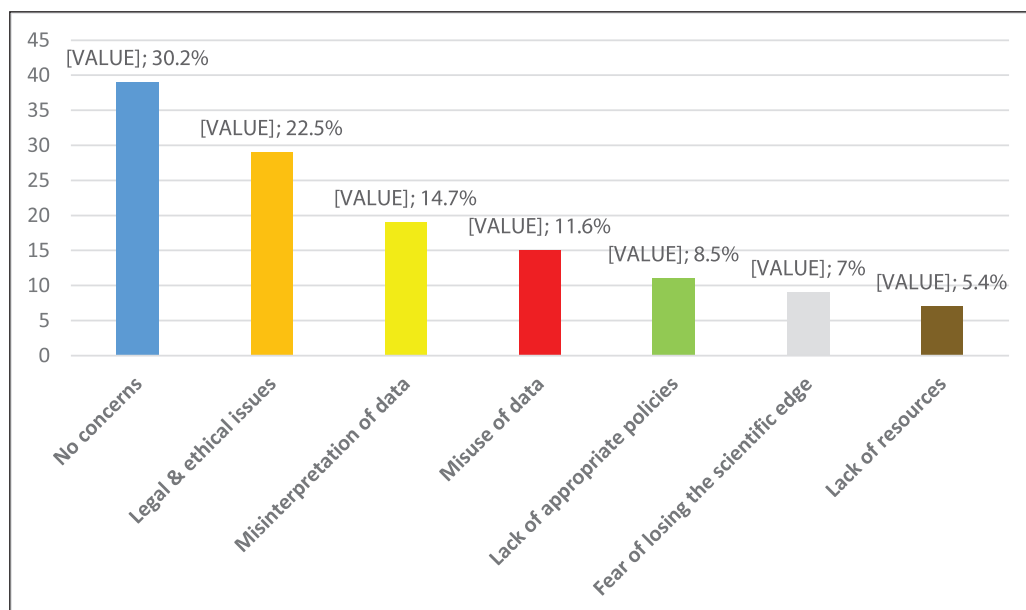


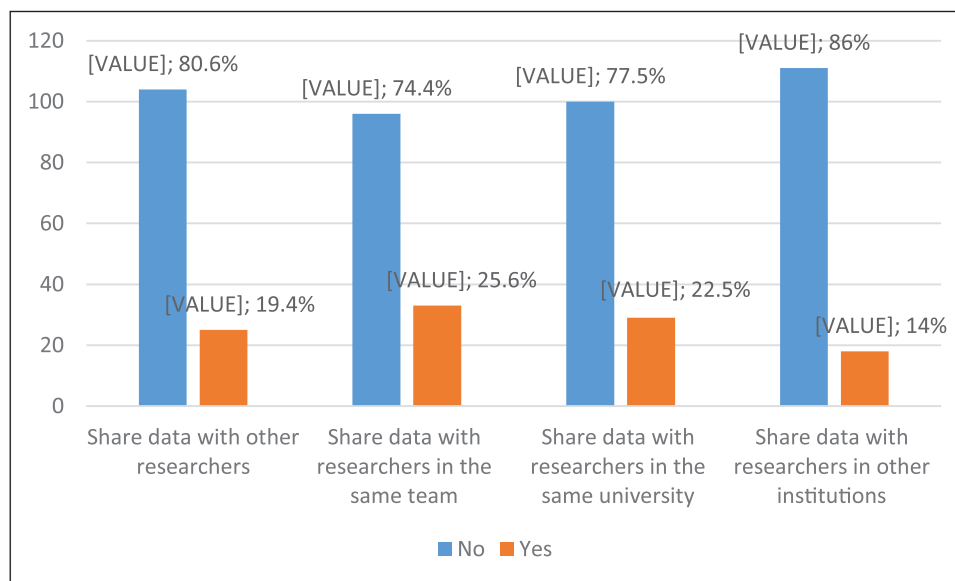**Figure 6:** Concerns about sharing data.

**Figure 7:** Current data sharing practices.

## 8. Study limitations

The biggest limitation of this study stems from respondents' characteristics. Due to under-representation of certain disciplines it proved difficult to compare the results by faculty. The over-representation of postgraduate students and emerging researchers also means that the results are more a reflection of their views than those of other researchers.

## 9. Conclusion and recommendations

From the findings of this study, it can be concluded that the majority of emerging researchers in South Africa preferred to use data from other sources and were not very keen on sharing their own data. While literature reviewed for this study indicate that a large number of researchers in institutions of other countries such as the Norwegian University of Life Sciences were open to the culture of sharing data, the same can hardly be said about the South African researchers in this study probably because most of them were early career researchers. Only 19.4% of South African researchers surveyed indicated that they currently share their research data with others. Further, researchers in South Africa expressed concerns in data sharing ranging from ethical issues, misuse of data to lack of resources. Ethical concerns, therefore, is one of the major barriers to data sharing among South African researchers. It can also be concluded that in terms of collaboration, emerging researchers in South Africa prefer to share their data with colleagues in their research environment as opposed to collaborate with researchers at other universities.

The results of this study have implications for libraries, research funders, universities and researchers in South Africa and other countries. These results point to low levels of data sharing culture and practices among South African researchers. Given that data management and sharing practices are in infancy in South Africa (Onyancha, 2016; Chiware and Becker 2018), these results point to a need for those involved in the promotion of data management and sharing practices to raise awareness about the importance of sharing data to researchers. These results challenge libraries to provide platforms that would assist researchers to store and share data on open data repositories. In the long run, it is hoped that this will change the data sharing culture of researchers in the country.

As a suggestion, funder mandates should be enforced as a way to ensure that there is compliance in terms of making research data available in open repositories for reuse. All researchers should be encouraged to submit their data management and data sharing plans at the proposal stage of applying for funds. In addition, incentives should be made available to researchers not just by the South Africa Department of Higher Education and Training, but also by universities as a way to stimulate the culture of data sharing. As an incentive, DHET may add an extra point to the publication subsidy to the published work whose data is shared in an open repository. Universities may consider researchers who share their data in open repositories for future promotions. In that way, there will be motivation for the researchers to share their data. Lastly,

in promoting research data management practices, libraries should explain institutional policies addressing ethical issues and data misuse to researchers in order to guide them accordingly. This will go a long way in allaying their fears and stimulate growth in open data practices. Due to the general low response rates to questionnaires, studies that utilise other methods such as in-depth and focus group interviews may provide more insights to the data management and sharing practices of researchers in South Africa.

## Data Accessibility Statement
The processed data set in the form of graphs used to write this paper can be found at: https://doi.org/10.25388/nwu.7381082.

## Additional File
The additional file for this article can be found as follows:

- **Appendix A.** ZA Digital literacy and Research Data Management Survey. DOI: https://doi.org/10.5334/dsj-2019-028.s1

## Acknowledgements

## Competing Interests
The authors have no competing interests to declare.

## References
**Bezuidenhout, L** and **Chakauya, E.** 2018. Hidden concerns of sharing research data by low/middle-income country scientists. *Global Bioethics*, 29(1): 39–54. DOI: https://doi.org/10.1080/11287462.2018.1441780

**Borgman, CL.** 2015. Big data, little data, no data. Cambridge: MIT Press. DOI: https://doi.org/10.7551/mitpress/9963.001.0001

**Buys, CM** and **Shaw, PL.** 2015. Data Management Practices Across an Institution: Survey and Report. *Journal of Librarianship and Scholarly Communication*, 3(2): eP1225. DOI: https://doi.org/10.7710/2162-3309.1225

**Chen, X** and **Wu, M.** 2017. Survey on the needs for chemistry research data management and sharing. *The Journal of Academic Librarianship*, 43(4): 346–353. DOI: https://doi.org/10.1016/j.acalib.2017.06.006

**Chiware, E** and **Becker, D.** 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.

**Cleary, M, Jackson, D** and **Walter, G.** 2013. Editorial. Research data ownership and dissemination: is it too simple to suggest that 'possession is nine-tenths of the law'? *Journal Of Clinical Nursing*, 22(15), 2087–2089. DOI: https://doi.org/10.1111/jocn.12140

**Dietrich, S, Van der Ham, J, Pras, A, Van Rijswijk-Deij, R, Shou, D, Sperotto, A, Van Wynsberghe, A** and **Zuck, LD.** 2014. Ethics in data sharing: developing a model for best practice. *IEEE Security and Privacy Workshops*. DOI: https://doi.org/10.1109/SPW.2014.43

**Elsayed, AM** and **Saleh, EI.** 2018. Research data management and sharing among researchers in Arab universities: an exploratory study. *IFLA Journal*. DOI: https://doi.org/10.1177/0340035218785196

**Elsevier.** 2018. Sharing research data. https://www.elsevier.com/authors/author-services/research-data.

**Engineering and Physical Sciences Research Council.** 2018. Scope and benefits. https://epsrc.ukri.org/about/standards/researchdata/scope/.

**European Commission.** 2012. Key Data on Education in Europe 2012 – Final Report – European. https://ec.europa.eu/eurostat/documents/3217494/5741409/978-92-9201-242-7-EN.PDF/d0dcb0da-5c52-4b33-becb-027f05e1651f.

**Howie, CT, Neethling, J** and **Louw, A.** 2006. Privacy and data protection. Pretoria: South African Law Reform Commission. https://www.gov.za/sites/default/files/gcis_document/201409/a25-02.pdf.

**JISC.** 2016. Research data assessment support Findings of the 2016 data assessment framework (DAF) surveys. https://www.researchgate.net/publication/318440591_Findings_of_the_2016_data_assessment_framework_DAF_surveys.

**Koopman, MM** and **De Jager, K.** 2016. Archiving South African digital research data: how ready are we? *South African Journal of Science*, 112(7/8): 1–7. DOI: https://doi.org/10.17159/sajs.2016/20150316

**Kvale, LH.** 2012. Data Sharing in the life sciences: a study of researchers at the Norwegian University of Life Sciences. (Masters thesis). Oslo: Universita Degli Studi Di Parma.

**Maluleka, JR** and **Onyancha, OB.** 2016. Research collaboration among Library and Information Science schools in South Africa (1991–2012): an informetrics study. *Mousaion*, 34(3): 36–59.

**Martone, ME, Garcia-Castro, A** and **Van den Bos, GR.** 2018. Data sharing in psychology. *American Psychologist*, 73(2), 111–125. DOI: https://doi.org/10.1037/amp0000242

**Michener, WK.** 2015. Ecological data sharing. *Ecological Informatics*, 29(1): 33–44. DOI: https://doi.org/10.1016/j.ecoinf.2015.06.010

**National Research Foundation.** 2015. Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research. http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded.

**North-West University.** 2018. Research ethics policy. http://www.nwu.ac.za/sites/www.nwu.ac.za/files/files/i-governance-management/policy/9P-9._Research%20Ethics%20Policy_e.pdf.

**OECD.** See Organisation for Economic Co-operation and Development.

**Onyancha, OB.** 2016. Open research data in Sub-Saharan Africa: a bibliometric study using the Data Citation Index. *Publishing Research Quarterly*, 32(3): 227–246. DOI: https://doi.org/10.1007/s12109-016-9463-6

**Organisation for Economic Co-operation and Development.** 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. http://www.oecd.org/sti/inno/38500813.pdf.

**Parker, M.** 2015. Exploring the Ethical Imperative for Data Sharing. In: O'Connell, ME and Plewes, TJ. *Sharing Research Data to Improve Public Health in Africa: A Workshop Summary*. Washington, DC: National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK321547/pdf/Bookshelf_NBK321547.pdf.

**Patterton, L, Bothma, TJD** and **Van Deventer, MJ.** 2018. From planning to practice: An action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science*, 84(2): 14–26. DOI: https://doi.org/10.7553/84-2-1761

**Renaut, S, Budden, AE, Gravel, D, Poisot, D** and **Peres-Neto, P.** 2018. Data management, archiving, and sharing for biologists and the role of research institutions in the technology-oriented age. *BioScience*, 68(6): 400–411. DOI: https://doi.org/10.1093/biosci/biy038

**Research Data Alliance and The Committee on Data for Science and Technology.** 2016. Legal interoperability of research data: principles and implementation guidelines. *RDA-CODATA Legal Interoperability Interest Group*. http://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final2.pdf.

**Roos, A.** 2008. Personal data protection in New Zealand: lesson for South Africa? *Potchefstroom Electronic Law Journal*, 11(4): 61–109. http://www.scielo.org.za/pdf/pelj/v11n4/v11n4a04.pdf. DOI: https://doi.org/10.4314/pelj.v11i4.42243

**Ross, MW, Iguchi, MY** and **Panicker, S.** 2018. Ethical aspects of data sharing and research participant protections. *American Psychologist*, 73(2): 138–145. DOI: https://doi.org/10.1037/amp0000240

**Schöpfel, J** and **Prost, H.** 2016. Research data management in social sciences and humanities: A survey at the University of Lille (France). *LIBREAS*, 29: 97–112.

**Seto, B** and **Luo, J.** 2007. Biomedical data sharing, security and standards. *Data Science Journal*, 6(17): 54–57. DOI: https://doi.org/10.2481/dsj.6.OD54

**Sieber, JE.** 2005. Ethics of sharing scientific and technological data: a heuristic for coping with complexity & uncertainty. *Data Science Journal*, 4(22): 165–170. DOI: https://doi.org/10.2481/dsj.4.165

**South Africa.** 1976. Plant Breeders' Act, 1976. Pretoria: Government Communication and Information System. https://www.gov.za/sites/default/files/gcis_document/201504/act-15-1976.pdf.

**South Africa.** 1978. South African Copyright Act Pretoria: Government Communication and Information System. https://juta.co.za/media/filestore/2018/06/Copyright_Amendment_Bill_B_draft.pdf.

**South Africa.** 1998. National Research Foundation ACT 23 OF 1998. http://www.nrf.ac.za/sites/default/files/documents/NTFAct.pdf.

**South Africa.** 2000. The Promotion of Access to Information Act, 2000 (PAIA). http://www.justice.gov.za/legislation/acts/2000-002.pdf.

**South Africa.** 2002. Electronic Communications and Transactions Act, 2002. Pretoria: Government Communication and Information System. https://www.gov.za/sites/default/files/gcis_document/201409/a25-02.pdf.

**South Africa.** 2008. Trade Marks Act, No. 194 of 1993. Pretoria: Government Communication and Information System. https://iponline.cipc.co.za/Publications/Acts/Trade_Marks.pdf.

**South Africa.** 2013. The Protection of Personal Information Act, No. 4 of 2013. http://www.justice.gov.za/inforeg/docs/InfoRegSA-POPIA-act2013-004.pdf.

**Tenopir, C, Allard, S, Douglass, K, Aydinoglu, AU, Wu, L, Read, E, Manoff, M** and **Frame, M.** 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6): e21101. DOI: https://doi.org/10.1371/journal.pone.0021101

**Tenopir, C, Dalton, ED, Allard, S, Frame, M, Pjesivac, I** and **Birch, B.** 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*, 10(8): e0134826. DOI: https://doi.org/10.1371/journal.pone.0134826

**University of Johannesburg.** 2007. Code of academic and research ethics. https://www.uj.ac.za/research/Documents/policy/Code%20of%20Academic%20and%20Research%20Ethics.pdf.

**University of Pretoria.** 2017. Research data management policy. Pretoria: University of Pretoria. https://www.up.ac.za/media/shared/12/ZP_Files/research-data-management-policy_august-2018.zp161094.pdf.

**Van den Eynden, V, Knight, G, Vlad, A, Radler, B, Tenopir, C, Leon, D, Manista, F, Whitworth, J** and **Corti, L.** 2016. Towards open research: practices, experiences, barriers and opportunities. London: London School of Hygiene and Tropical Medicine and the UK Data Service. https://figshare.com/articles/Survey_of_Wellcome_researchers_and_their_attitudes_to_open_research/4055448/1.

**Wiley.** 2014. Research Data Sharing Insights. https://3spxpi1radr22mzge33bla91-wpengine.netdna-ssl.com/wp-content/uploads/2014/11/researcher-data-insights-infographic-final.pdf.

**Wilkinson, MD, Dumontier, M, Aalbersberg, IJJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJG, Groth, P, Goble, C, Grethe, JS, Heringa, J, Hoen, PAC't, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Rocca-Serra, P, Roos, M, van Schaik, R, Sansone, S-A, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, MA, Thompson, M, van der Lei, J, van Mulligen, E, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J** and **Mons, B.** 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data*, 3: 160018. https://www.nature.com/articles/sdata201618.pdf. DOI: https://doi.org/10.1038/sdata.2016.18

**Wouters, P** and **Haak, W.** 2017. Open data: the researcher perspective. https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf.