

Data literacy training needs of researchers at South African universities

Data literacy
training needs

Mathew Moyo

*Library and Information Service, North-West University,
Potchefstroom, South Africa, and*

Siviwe Bangani

*Library and Information Service, Stellenbosch University,
Stellenbosch, South Africa*

Received 5 February 2023
Revised 10 April 2023
Accepted 22 May 2023

Abstract

Purpose – The aim of this study was to determine data literacy (DL) training needs of researchers at South African public universities. The outcome of this study would assist librarians and researchers in developing a DL training programme which addressed identified needs.

Design/methodology/approach – A survey research method was used to gather data from researchers at these universities by convenience. Online questionnaires were distributed to public universities through library directors for further distribution to researchers.

Findings – The results indicate low levels of DL training at the respondent South African public universities with most researchers indicating that they had not received any formal training on DL. A few researchers indicated that they would welcome DL training.

Research limitations/implications – This study was exploratory in nature and data was received from eight universities, which is not representative of all the 26 public universities in South Africa. Nonetheless, the low DL confirmed by the majority in the realised sample is indicative of the need to further investigate the subject.

Practical implications – Librarians and research support personnel should collaborate on the development of DL training courses, workshops and materials used by researchers at institutions of higher learning to enhance DLs on campus.

Originality/value – This study may be novel in South Africa in investigating the DL training needs of researchers at several universities and contributes to the growing body of literature on research data management

Keywords Data literacy, Research data management, Researchers, Training needs, University, South Africa

Paper type Research paper

© Mathew Moyo and Siviwe Bangani. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial & non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors acknowledge the ECIL Data Literacy Research Team, the developers of the online questionnaire (Appendix), for extending an invite to participate in the study and for allowing us to share the results of the survey. This paper was developed from a presentation prepared for the 2019 IFLA's Big Data Special Interest Group, Frankfurt, Germany.



1. Introduction

The United Nations Educational, Scientific and Cultural Organization (UNESCO) Institution for Statistics (2022) defines literacy as the ability to identify, understand, interpret, create, communicate and compute, using printed materials associated with varying contexts. The United Nations Educational, Scientific and Cultural Organisation (UNESCO) International Bureau of Education (2022) has since extended the traditional definition of literacy to encompass other literacies such as linguistic, visual, audio, spatial and gestural literacies hence the reference to multiliteracies.

According to Onyancha (2020, p. 117), there are various types of literacies, and these include reading and writing literacy, information literacy, digital literacy, academic literacy, financial literacy, physical literacy and data literacy (DL). Onyancha lists 42 literacies although he concedes that the list is not exhaustive but only inclusive of literacies associated with information literacy. Although most of these forms of literacies have largely been explored, DL is a new and largely unexplored form of literacy (Mandinach and Gummer, 2016; Patterton *et al.*, 2018), particularly in South Africa (Chiwere, 2020; Chiwere and Becker, 2018). However, it is as important as any other form of literacy (Muronga and Ogunlaja, 2022).

DL comprises competencies required by researchers to work with data (Schneider, 2013 as cited by Vilar and Zabukovec, 2018). It is the ability to comprehend and critically evaluate data in all aspects of life (Fontichiaro and Oehrli, 2016). Critical skills expected of data-literate individuals include the ability to collect and organise, analyse and summarise and synthesise and prioritise data in an ethical manner (Burrell *et al.*, 2020; Fontichiaro and Oehrli, 2016; Gummer and Mandinach, 2015). Kubovics and Zaušková (2020, p. 343) describe DL as recognition of the need for data, creation of one's data or search or acquisition of already created data, critical assessment of data and their sources, data management, their storage including long-term archiving, data sharing including open access and use of data as well as ethical and legal aspects. The above process of DL situates it in the information literacy paradigm which also begins with the recognition of the information need and ends with the ethical and legal use of information (Onyancha, 2018, p. 116; Koltay, 2017; Vilar and Zabukovec, 2018). Schneider (2013) also noted parallels between information literacy and DL referring to the latter as the offspring of the former. Meanwhile, as early as 2004, Schield had sought to show interrelatedness between DL, information literacy and statistical literacy. The author postulated that, among the three literacies, librarians will find it difficult to promote one type of literacy over others meaning that they will have to find a way to teach all the three literacies either parallel or together. This is because of the fact that a great deal of information requires statistical and data knowledge.

Like all literacies, DL is transferrable through conscious skills transfer and imparting of knowledge. Research-supporting departments such as libraries (and e-braries) should develop mechanisms which ensure that their services remain relevant in a fast-changing research landscape. These include support for the DL initiatives of the universities (Chiwere, 2020). In that regard, Calzada-Prado and Marzal (2013) validate Chawinga and Zinn (2019) in the view that academic libraries are well-placed to use their experience, skills and expertise in delivering information literacy to teach DL skills.

However, before the current study, it was not clear what the views of South African researchers regarding DL training were, and whether they aligned with those of Malawi and elsewhere as reported by Chawinga (2019), Koltay (2017), Vilar and Zabukovec (2018) and many others. Previous studies conducted in South Africa either focussed on one institution (Patterton; Patterton *et al.*, 2018) or libraries/librarians (Chiwere and Becker, 2018; Kahn *et al.*, 2014). This study strove to find out the DL training needs of researchers at South African public universities.

2. Objectives of the study

In particular, the study sought to answer the objectives designed to:

- determine the attendance of formal DL training by researchers in South African public universities;
- establish the aspects of DL training received by the researchers;
- verify the views of the researchers regarding the necessity for DL training; and
- ascertain the DL training requirements of the researchers.

3. Literature review

The literature review highlights both the South African and the international perspectives on DL focusing particularly on researchers' DL needs. Globally, the driving force behind DL is the requirement by some journals, governments and funding agencies that research data must be shared openly and widely (Bangani and Moyo, 2019; Chiware and Becker, 2018; Vilar and Zabukovec, 2018). According to Carlson *et al.* (2011, p. 2), there are other underlying factors which pushed interest in DL to the fore. These are:

- “the capacity to store massive amounts of data”;
- “a robust and growing suite of advanced informational and computational data analysis”; and
- increased capacity for data visualisation using advanced computational tools.

In the past, DL was associated with researchers in the pure sciences, but other disciplines such as humanities and social sciences have come on board of late (Koltay, 2015).

In South Africa, DL is gaining momentum as a result of the push by the South African Government through the Department of Higher Education and Training, Science and Technology, and the National Research Foundation (Onyancha, 2018; Bangani and Moyo, 2019). These departments lead the organisations calling on researchers to manage and share data widely. To ensure compliance, policies and statements in support of open data have been developed and issued in the country (Bangani and Moyo, 2019).

In addition, Universities South Africa (USAf), a body involving all public universities in South Africa, has been engaged in awareness creation workshops for researchers in the higher education sector [Universities South Africa (USAf), 2015]. These workshops were organised and conducted jointly with government departments. The aim was to ensure that researchers need to be data literate and take steps to develop sufficient DL skills to operate in a data-intensive environment. At the time of writing this article, a draft Open Science Policy for South Africa had been circulated for public comments, with timeframes, for its approval already envisaged (Universities South Africa, 2022).

However, questions arise about whether researchers have the necessary skills to manage the full cycle of data. The South African–European Union (SA-EU) Open Science Dialogue report notes that “upskilling interventions across domains: educators, academics, researchers, etc., hold the key for success” (South African–European Union, 2018, p. 24). Chawinga (2019) agrees but adds that most researchers in the universities in Malawi yearn for more DL training opportunities. Similar to Chawinga, Koltay (2017) and Vilar and Zabukovec (2018) identify DL as an essential skill for researchers. Koltay (2017, pp. 10–11) emphasises this by stating that “acquiring data literacy skills is thus an issue for researchers, including graduate and doctoral students, who need to become data literate”.

3.1 Formal data literacy training attendance by researchers

There is consensus in literature that DL skills are critical for researchers (Chawinga and Zimm, 2019; Schneider, 2013), librarians (Chiware and Becker, 2018) and students (Fontichiaro and Oehrli, 2016; Stephenson and Caravello, 2007). Attendance of formal DL training has been the focus of various studies in South Africa (Chiware and Becker, 2018; Kahn *et al.*, 2014; Patterton, 2016; Patterton *et al.*, 2018) and many other countries (Chawinga, 2019; Elsayed and Saleh, 2018; Federer *et al.*, 2016; Koltay, 2017; Calzada-Prado and Marzal, 2013; Vilar and Zabukovec, 2018). These studies either treat the subject from a disciplinary perspective (Federer *et al.*, 2016) or from multidisciplinary perspectives (Chawinga, 2019).

Previously, researchers relied on trial and error and on-the-job experience to develop and polish their DL skills. As an example, Koltay (2017) explored DL for researchers and data librarians. Koltay's study demonstrates that researchers are learning data management and curation on the job and in an ad hoc fashion, and that they are not satisfied with their level of DL. Federer *et al.*'s (2016) study focused on the imparting of DL skills to biomedical researchers in the USA. The researchers found that 77% of biomedical researchers have not attended any form of DL training. This finding is similar to studies in other countries. For example, a Slovenian study by Vilar and Zabukovec (2018) clearly shows that the majority of researchers who responded to their survey had not attended any form of DL training. The majority (54%) expressed a desire to attend such formal training if it were to be offered.

In addition, Chawinga's (2019) study sought to explore the RDM practices of universities at Malawi's universities using a mixed-methods approach. Regarding DL training, Chawinga (2019) found that there was a general failure of universities in that country to organise RDM or DL workshops which led to only 24.2% of respondents having attended any formal DL training. However, the majority of respondents were willing to attend formal DL skills.

In South Africa, Patterton *et al.* (2018) found that 88% of emerging researchers at the Council for Scientific and Industrial Research never attended any DL training. The authors concluded that because of this lack of training there were skills gaps that could be addressed through DL training.

3.2 Aspects of data literacy training received by researchers

Scholars differ on what constitutes DL skills and even the terminology assigned to such DL skills (Koltay, 2017). The emphasis on certain skills may differ depending on disciplinary variations and differences in roles. Despite this, aspects of DL training already received by researchers are of interest to data scholars. Relevant literature in this area includes Al-Jaradat (2021), Chawinga (2019), Elsayed and Saleh (2018), Federer *et al.* (2016), Patterton (2016), Kahn *et al.* (2014) and Vilar and Zabukovec (2018).

Regarding DL skills, Carlson *et al.* (2011) provide an expansive list involving metadata standards, standardising documentation processes, maintaining relationships among data master files and versioning, ethics, basic database skills, quality assurance and preservation. Al-Jaradat (2021) mentions metadata/cataloguing, networking/collaboration skills, data preservation techniques and tools/systems/software. Federer *et al.* (2016) list metadata, ontology, collaboration, data mining, reuse, visualisation, retention, deposit and RDM. Similar to this study, Majid *et al.* (2018) and Vilar and Zabukovec (2018) list data management plans (DMPs), metadata, consistent file naming, version control of data sets and data citation styles. Chawinga (2019, p. 221) lists metadata, hardware troubleshooting, data appraisal, data management, data retrieval, curation life cycle, preservation strategies, data citation, data transformation and hardware and software installation.

Despite differences on what should constitute DL skills, this has not stopped scholars from assessing DL skills training already received by researchers. Vilar and Zabukovec (2018)

assessed the DL skills of researchers in Slovenia in five aspects: metadata, version control of data sets, writing of DMPs, consistent file naming and data citation. The study found that around 17.9% of respondents received data citation training, 11% received metadata training, 5% in DMP and only 2% received training on version control of data sets. A similar study conducted by researchers from Nanyang Technological University in Singapore found that 18.7% of respondents attended DMP training, 5.8% attended metadata training, 7.1% attended consistent file naming, 6.6% attended version control of data sets and 26.1% attended data citation styles.

[Chawinga \(2019\)](#) assessed research DL skills already received by researchers in Malawi and verified that only 24.2% had attended any type of training. Only 7.4% of the researchers indicated that they had received training on metadata, 24.1% received training on DMP and only 4.8% received training on migrating data to newer file formats. In a South African study of emerging researchers at the Council for Scientific and Industrial Research (CSIR), [Patterton \(2016\)](#) confirmed that only 4% of the researchers received any form of DL training while 8% were not sure whether they ever received training.

3.3 Views of researchers regarding the necessity for data literacy training

Researchers proffer different views about whether or not it is necessary to undergo DL training. Studies in this area include [Chawinga \(2019\)](#), [Elsayed and Saleh \(2018\)](#), [Federer et al. \(2016\)](#), [Patterton \(2016\)](#), and [Vilar and Zabukovec \(2018\)](#).

In studying the DL training needs of biomedical researchers affiliated with the National Institute of Health (NIH), [Federer et al. \(2016\)](#) speculated that because of the high proportion of NIH's biomedical researchers who lacked certain DL skills, there is need for DL training in that institute. However, it was not clear if the researchers' views of the DL training were canvassed. Elsewhere, some studies show that having a high proportion of researchers who do not have sufficient DL skills does not always translate to researchers requiring the relevant training.

In [Vilar and Zabukovec \(2018\)](#), for example, most researchers indicated that they did not need metadata skills training despite a paucity in those skills. [Koltay \(2017\)](#) identifies metadata skills training as a skill that is often exclusive to librarians, meaning researchers do not always see a need for metadata training skills.

In the Arab countries, [Elsayed and Saleh \(2018\)](#) found that close to 57% of respondents in three universities from Egypt, Jordan and Saudi Arabia would desire DL skills training on various aspects of RDM. [Patterton \(2016\)](#) regarded the high number of researchers who indicated that they do not need DL training in her study as alarming. In [Patterton's](#) study, 21% of researchers indicated that they do need to undergo DL training despite lacking in those skills.

The differences between [Elsayed and Saleh \(2018\)](#), [Koltay \(2017\)](#), [Vilar and Zabukovec \(2018\)](#), [Patterton \(2016\)](#) and [Chawinga's \(2019\)](#) studies point to a possibility that there are different views by institutions on the necessity for DL training.

3.4 Data literacy training requirements of researchers

Despite many researchers expressing negative views on the need for DL training, when requested to identify training requirements, they are often able to do so. There are often differences in what the researchers indicate as their training requirements.

In [Majid et al.'s \(2018\)](#) study, 62.2% of the respondents expressed that they would be interested to attend training on DMPs, 57.3% would like to attend training on metadata, 43.6% on version control of data sets, 42% on consistent file naming and 39.8% on data citation styles. According to [Chawinga \(2019\)](#), in one Malawian university, 68% of researchers

expressed that they required training in writing DMPs. This number was higher in another university at 85.7%. Regarding metadata, 94% of researchers at one university expressed the necessity for metadata training and the percentage stood at 91.3% in another institution.

In a South African study by [Patterton \(2016\)](#), 58% of emerging researchers from the CSIR indicated that they would like training on DMPs, 42% on metadata creation and 50% would like DL training on data citation.

However, despite what appears to be a plurality of perspectives in the area of DL, available studies either did not have a wide focus or were concentrated on librarians rather than researchers. [Patterton \(2016\)](#) and [Patterton *et al.* \(2018\)](#), for example, focused on the CSIR. [Chiwane and Becker \(2018\)](#) focused on libraries in Southern Africa and [Kahn *et al.* \(2014\)](#) focused on librarians in South Africa. These studies' focus was RDM initiatives generally, with DL covered as part of a plethora of RDM activities while this study focuses on the issue of DL training specifically. This study has a wider reach in this area than previous South African studies mentioned above.

4. Methodology

This study was conducted as part of a multinational survey which was initiated by a DL Research Team comprising members from universities in England, France and Turkey. The researchers agreed to represent South Africa following an invitation received from the originators of the survey. As such, the researchers used a pre-prepared online questionnaire guide, included in this study as [Appendix](#), to determine the DL needs of researchers in South Africa. Twenty-six online questionnaires were distributed to 26 public universities in South Africa through the respective library directors between 2018 and 2019. The library directors were requested to further distribute the questionnaires to the target participants of the survey who were academics, researchers, masters' and doctoral students at public universities in South Africa. Although some participants chose not to indicate their institutions of affiliation, noted responses were received from: Cape Peninsula University of Technology, North-West University (NWU), University of Venda, Sol Plaatjie University, Tshwane University of Technology and the University of South Africa. In total, 141 responses were received, of which 140 were found to be usable. Although the response rate was lower than anticipated, the researchers decided to use the data in line with previous studies in the area ([Bangani and Moyo, 2019](#)). For example, [Vilar and Zabukovec's \(2018\)](#) response rate in Slovenia was exactly the same as that of this study. The questions contained single-response and multiple-response multiple-choice questions. All responses were sent directly to the original creators of the survey through a central email address. The researchers then requested to get the South African data which was received in an Excel format. This study focused on the responses to four questions that related to the DL training needs of the researchers. This focus assisted the researchers to respond to a call for papers from IFLA's 2019 Big Data Special Interest group pre-conference where the preliminary results of the study were shared. Although biographical information and other researcher categorisation information was gathered for the entire project, it was deemed unnecessary for this part of the study because the questions focussed on funder mandate issues. For instance, all researchers receiving government funding in South Africa are required to share the data used in their studies or research through the institutional data repository, while research students are required to have a DMP. It was therefore deemed that the DL requirements for all researcher categories were the same.

Ethical clearance was requested and approved by the NWU Ethics Committee. Thereafter, a letter of request with a link to the online survey was emailed 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 according to their ethical policies, and the researchers complied with such policies where applicable. Other directors simply proceeded to distribute the questionnaires without asking for further compliance particularly given that the NWU ethical clearance was attached.

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

5. Study findings

The findings are organised according to the themes derived from the objectives of the study as discussed below:

5.1 Attendance of formal data literacy training by the researchers

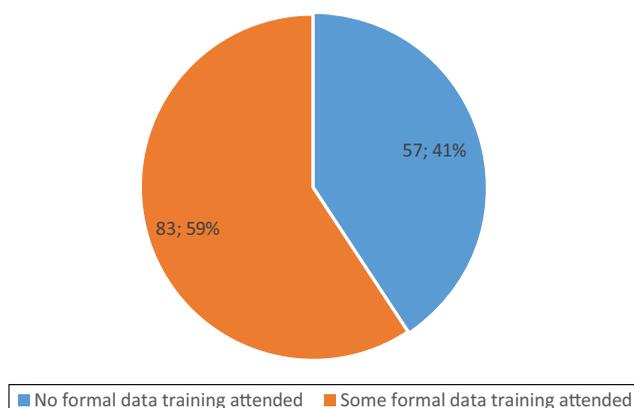
To determine the attendance of previous DL training, respondents were asked to indicate whether they had attended any formal DL training before and given two options:

- (1) to indicate whether they had not attended any DL training; and
- (2) to indicate whether they had attended some formal data training before.

The results in [Figure 1](#) show that the majority of the respondents (59% or 83 out of 140) had attended some formal data training before, whereas 57 (or 41%) of all the respondents never attended any formal DL training.

5.2 Data literacy training received

The purpose here was to find out the aspects of DL training that the participants had received. Participants were asked to indicate whether they had received training on pre-selected aspects



Source: Figure by authors

Figure 1.
Previous attendance of any formal data literacy training ($N = 140$)

of DL. The results in Figure 2 indicate that 7% (10 out of 140) of the participants had received training in DMPs, meaning that close to 93% (130 out of 140) of all the participants indicated that they had not received any training on this critical aspect of RDM. It was more concerning that the frequency was even lower when it comes to metadata training. A negligible 2% (3 out of 140) of the participants stated that they had received any form of metadata training. Close to 3% (4 out of 140) had received training on version control of data sets. This figure was slightly better than for consistent file naming; 5% (7 out of 140) of all the participants indicated that they had received training on this aspect against 133 who had not. The last aspect of the question required participants to indicate whether they had received any formal training on data citation styles. This area fared better with close to 18% (25 out of 140) participants indicating that they had received training on this matter while 115 had not.

5.3 Necessity to have formal data literacy training

This question sought to determine the respondents' views about the significance of formal DL training. This was to respond to the third question which sought to verify the views of the researchers regarding the necessity for DL training. As indicated in Figure 3, only 2% (3 out of 140) of the respondents agreed that it is important to have formal DL training. However, a staggering 98% (137 out of 140) of participants indicated that it was not necessary to have formal training. Despite the researchers not having received a lot of formal training as shown in Figure 2, very few would be willing to attend formal training.

5.4 Data literacy training requirements of the researchers

This theme sought to find out the researchers' DL training requirements. Respondents chose answers from the following five DL training areas:

- DMP;
- metadata;
- version control of data sets;
- consistency file naming; and
- data citation styles.

The results in Figure 4 show that 82 of the 140 (58.6%) respondents indicated that they needed training on DMPs. Sixty-eight (48.6% of 140) of the respondents indicated that they

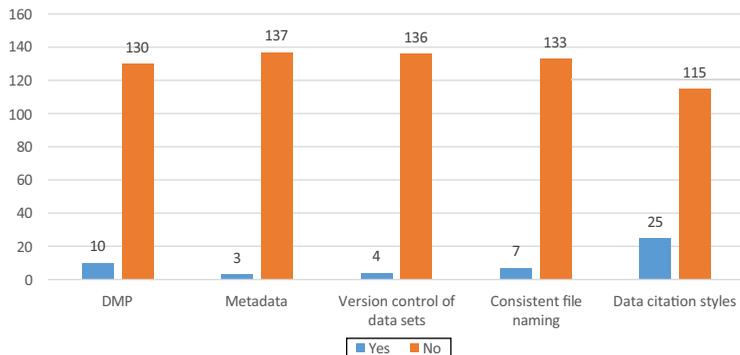
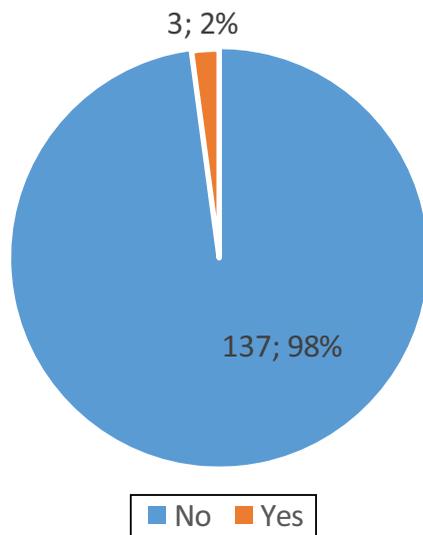


Figure 2. Previous formal data literacy training received (N = 140)

Source: Figure by authors

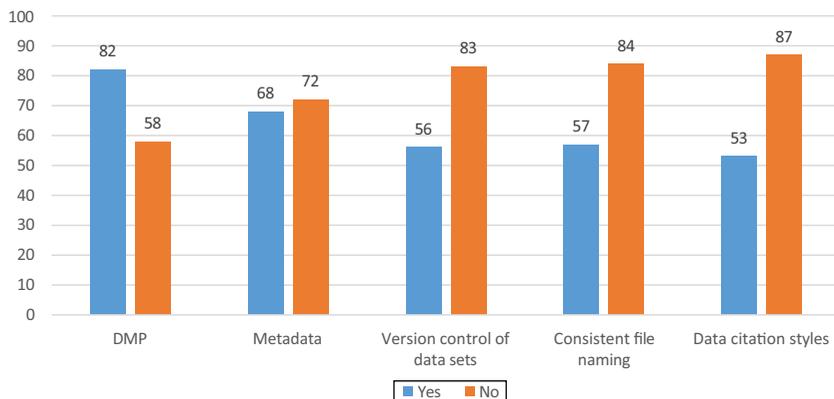


Data literacy training needs

Source: Figure by authors

Figure 3. Necessity of formal data literacy training

needed training on metadata. Regarding version control of data sets, 56 (40% of 140) of the respondents conceded that they needed training on the aspect, while the majority indicated that they would not need training on this theme (recast). Regarding consistency file naming, 57 (40.7% of 140) participants said they required training on the aspect. The last question on DL training requirements of researchers relates to data citation styles. Of the 140 participants, 53 (37.9% of 140) indicated that they had training requirements on the theme, whereas 87 respondents saw no need for training on the item. This figure therefore indicates that in providing training, librarians should put more emphasis in DMP and metadata training because these are priority areas identified by the researchers.



Source: Figure by authors

Figure 4. Data literacy training requirements of researchers (N = 140)

6. Discussion of the results

The discussion of results is aligned to the objectives of the study and the findings.

6.1 Attendance of formal data literacy training by the researchers

The study showed that the majority of the researchers have previously attended some form of formal DL training. The study involved academic researchers and doctoral students who might have received data training in one way or the other during the course of their work or studies. This justifies the high number of respondents (59%) who indicated that they had attended some formal data training.

The results of this study rebutted those of earlier studies by [Chawinga \(2019\)](#), [Koltay \(2017\)](#), [Majid et al. \(2018\)](#), [Patterton et al. \(2018\)](#) and [Vilar and Zabukovec \(2018\)](#). Unlike [Koltay \(2017\)](#), the majority of this study's respondents stated that they have received formal training meaning that they did not have to rely on ad hoc on-the-job forms of DL skills acquisition. Previous studies showed that the majority of the respondents had not received formal training. In [Chawinga \(2019\)](#), 74.2% of the researchers had not attended formal training. The frequency was 77% in [Federer et al. \(2016\)](#), 60.2% in [Majid et al. \(2018\)](#) and 88% in [Patterton et al. \(2018\)](#).

Previous studies ([Chiware and Becker, 2018](#)) and [Onyancha, 2018](#)) noted that librarians were already offering some form of formal training in research DL. The current study's results might be an affirmation that these interventions are beginning to bear fruits. In this study, more researchers indicated that they received some form of DL training.

However, the high number (57 or 41%) of those who indicated that they had not attended any formal data training should be concerning. It suggests that librarians and research support offices and entities providing training to researchers still have a long way to go in ensuring that researchers at all levels become data literate. This is in line with [Chiware's \(2020\)](#) assertion that data librarianship is still a growing area that has not yet reached maturity in Southern Africa and globally.

6.2 Data literacy training received

The figures for researchers who received formal training were much lower than those who did not receive formal training. There were more researchers who had received formal training on data citation styles than all the other aspects followed by DMP training and consistent file naming with version control of data sets and metadata coming second last and last, respectively.

The foregoing finding contradicts the findings (above) presented earlier where most researchers stated that they had received some form of formal training. However, it must be borne in mind that the list of DL training options provided for the researchers to select from in this study is by no means exhaustive. The researchers may have been trained on other aspects that appear either on [Lyon et al.'s \(2011\)](#) or [Chawinga's \(2019\)](#) more exhaustive lists of DL skills and possible interventions.

The current results compare relatively well with those of similar studies by [Majid et al. \(2018\)](#) and [Vilar and Zabukovec \(2018\)](#). In the case of [Majid et al. \(2018\)](#), the training requirements are similarly placed except for version control of data sets and consistent file naming which came in third and fourth places, respectively. In this study, consistent file naming was placed third in terms of training requirements while version control of data set at number four. The difference between [Vilar and Zabukovec \(2018\)](#) and this study is in metadata training and consistent file naming. Coincidentally, around 25 out of 140 respondents in [Vilar and Zabukovec \(2018\)](#) also indicated that they had received training on data citation. However, in [Vilar and Zabukovec \(2018\)](#), metadata came second, with around

11% of all researchers indicating that they received metadata training. This figure stood at 2% in this study. Another aspect where [Vilar and Zabukovec \(2018\)](#) compare well with this study is DMP training. Around 5% of all respondents in [Vilar and Zabukovec \(2018\)](#) indicated that they received DMP training, while this figure stands at 7% in this study. On consistent file naming, this study out-performed [Vilar and Zabukovec \(2018\)](#) whose figure was around 3 out of 140. Regarding version control of data sets, the results of this study confirm and validate those of [Vilar and Zabukovec \(2018\)](#).

The higher number of researchers who stated that they received training on data citation styles should not come as a surprise. Many academic libraries used the ability for researchers to accumulate citations from their published data as a way of marketing library data services. [Patterton et al. \(2018, p. 22\)](#) confess that “researchers are more easily persuaded to add their data to a repository when they know the data would be cited”.

Interestingly, all the ten respondents who claimed to have had previous formal training on DMPs indicated that they would still prefer to attend the training in the future. This may suggest that the formal training they had was not effective enough hence the desire for a follow-up training. Related to this interesting aspect is also the fact that three respondents who claimed to have had training on metadata would still like to receive training on the same. This may also suggest that the respondents never used the knowledge gained during the training session and as such, retraining was required. The trend was the same for four of the seven respondents who claimed to have had training on consistent file naming. The researchers indicated that they would like to repeat the same training. The same goes with three of the four respondents who indicated that they have had previous training on version control of data. Furthermore, 10 of the 25 respondents who indicated that they have had previous training in data citation styles would like a refresher course on the same. Overall, a huge need for DL is portrayed by the responses as shown in [Figure 2](#).

6.3 Necessity for formal data literacy training

Regarding the necessity for formal DL training for researchers, a disparity can be noticed between the training received as stated in Subsection 6.2, training desired in Subsection 6.4 and the necessity for formal DL training. Although most of the researchers stated that they had not received training in the specific aspects listed in Subsection 6.2, they went on to state that they do not see formal DL training as necessary.

[Vilar and Zabukovec \(2018\)](#) also faced a similar conundrum when most researchers in their study indicated that they lack metadata skills but were not interested in relevant training. Earlier, [Patterton \(2016\)](#) was alarmed to notice that 21% of CSIR researchers were not interested in DL skills training. It is possible that although the researchers recognise their shortcomings, they do not feel that it is their responsibility to be trained in those aspects, preferring somebody else (e.g. libraries, IT and other academic support functions) to do these for them. Perhaps, [Carlson et al.'s \(2011\)](#) interviews with faculty provide some clues to this. In [Carlson et al. \(2011\)](#), researchers appeared to differentiate some data management aspects from the rest of their research, preferring to outsource them to graduate students. Some researchers felt that DL training was required for the students but not themselves and this training would have to be done by someone else. [Chawinga \(2019\)](#) found related sentiments that most researchers prefer to consult librarians for their data management needs, further giving credence to the possibility that researchers would prefer somebody else to handle some aspects of data.

From the researchers' experiences with equivalent personnel at the university level, it is also possible that the respondents misconstrued this question to refer to a course that they would be obliged to attend at the expense of their other research duties. It is also our

experience that if a training intervention is not presented in a formal way, chances are that most of the intended beneficiaries will not partake in it, or the training will be ignored. In the South African context, there is also the issue of credits assigned to formal learning programmes, and the concern always is that there is no room for new credits. All these issues may have influenced the respondents' choice of answer.

6.4 Data literacy training requirements of the researchers

Despite indicating that they do not see any necessity for DL training, the researchers were asked for their DL training requirements. The majority needed training in DMPs followed by metadata, consistent file naming, version control of data assets and data citation styles. It was encouraging that researchers recognised their weaknesses and were able to choose the areas where they needed assistance for improvement. Earlier, in Subsection 6.2, it was suggested that researchers most lacked training in metadata, version control of data sets and consistent file naming followed by DMP and data citation styles.

As hinted earlier, submission of a DMP is a requirement for many funding agencies (including the NRF) when researchers are applying for funding (Bangani and Moyo, 2019; Patterton *et al.*, 2018); therefore, the high number of respondents who indicated that they needed training on DMPs might attest to this requirement. Yet disagreement is evident between Federer *et al.* (2016) and the current study's results. While this study's respondents rated DMP training highly, the former's participants rated the need for DMP training as low in their work. However, this difference might be attributed to the time of Federer's study during which many funders did not have DMP mandates. Apparently, Federer *et al.* (2016) themselves commented that the low rank of DMP was set to likely change as funders start to make the writing of DMPs mandatory.

Historically, metadata work was associated with library and information professionals, and this could point to the reasons why a large number (72 out of 140) of the respondents indicated that they did not need any training on this aspect. Earlier statements of researchers preferring to outsource at least some aspects of data management may also be applicable. Koltay (2017) identifies metadata as one area that could be regarded as an exclusive task of librarians, although reading and interpreting metadata was identified as critical for researchers. This means that researchers would still need some form of metadata training.

These results contradict those of Chawinga (2019). In Chawinga (2019), 92.9% and 94% of researchers from two Malawian universities indicated that they would like training in metadata, whereas 68% and 85.7% needed DMP training. In this study, 58.6% of researchers required DMP training, whereas 48.6% required metadata training. However, it must be noted that Chawinga's (2019) respondents came from a low base. Only 24.2% of respondents in Chawinga indicated that they had attended any form of DL training as opposed to 59% in this study. It is not clear whether any of Chawinga's respondents had previously attended metadata and DMP training specifically, while a few researchers in this study stated that they had attended some training in those two areas. These results are a vindication of Patterton *et al.* (2018) who advocate for the provision of online DMP tools in libraries to guide researchers and minimise duplication of effort.

7. Conclusions and recommendations

This study strove to determine the DL training needs of researchers in South African public universities with a view to developing a DL training programme that would address the identified needs and training gaps. The results of the study exposed gaping holes in the DL training levels of South African researchers. These include a general lack of training in key aspects of data management such as writing DMPs, version control of data sets, ensuring

consistency in file naming and data citation styles. A few South African researchers attended training on all these aspects. Despite these limitations, reluctance to attend DL training was noticed. There is an apparent desire by most researchers to outsource some of the key data management functions. Furthermore, the results also pointed to DMPs and metadata training as priority areas for researchers in South Africa.

The study serves as a reminder that DL is no longer a choice but critical skill for researchers and librarians to function in a data-intensive environment. With advances in both technology and research as well as requirements from funders, ethical bodies and governments, users of the internet are generally expected to provide their own metadata for works which they post on the Web to enhance the discoverability of the information, as well as to develop own metadata and DMPs not only for research grant applications but also for research in general. Researchers who do not hold DL skills may struggle to cope in a data-intensive digital world.

These results serve as a wake-up call for a change of attitude towards RDM from academic libraries to adjust their DL training programs to the data practices of the research community. Furthermore, these results point to a need for libraries to engage in more advocacy work, raise awareness and intensify their DL skills training as endorsed by Vilar and Zabukovec (2018). Fontichiaro and Oehrli (2016) echoed Shied (2004) in identifying librarians as eminently placed to teach DL skills given their unique position as “cross-disciplinary pollinators” and generalists who do not only focus on a specific discipline. This can include creation and development of online content in the form of LibGuides, training videos and other online learning material, including the holding of webinars to complement the DL skills training. Given that DL and information literacy complement each other (Shied, 2004), libraries should treat DL with the same urgency as information literacy skills. Moreover, researchers can attend data training and learning opportunities provided by various universities in South Africa.

As recommendations, academic libraries that have not begun providing DL should start to do so as a matter of urgency to avoid the danger of being irrelevant in the data-intensive and digital age. Secondly, libraries should consider merging DL into aspects of their digital information literacy programmes as proposed by Burress *et al.* (2020). That way, librarians will save resources and time as they will simply market and advocate the programmes simultaneously. A disparity was noticed wherein researchers overwhelmingly rejected formal DL training despite an apparent need based on the low DL skills levels indicated. As a follow-up to the current study, it is recommended that qualitative studies that enable deeper understanding of experiences be undertaken to help shed some more light to this conundrum.

References

- Academic of Science South Africa (ASSAf) (2022), “Virtual stakeholder workshop to consult on the draft open science policy”, available at: www.assaf.org.za/files/2022/News%202022/Invitation%20%20Open%20Science20Stakeholder%20Workshop%202022.pdf (accessed 13 April 2022).
- Al-Jaradat, O.M. (2021), “Research data management (RDM) in Jordanian public university libraries: present status, challenges and future perspectives”, *The Journal of Academic Librarianship*, Vol. 47 No. 5, p. 102378.
- Bangani, S. and Moyo, M. (2019), “Data sharing practices among researchers at South African universities”, *Data Science Journal*, Vol. 18 No. 1, pp. 1-14.
- Burress, T., Mann, E. and Neville, T. (2020), “Exploring data literacy via a librarian-faculty learning community: a case study”, *The Journal of Academic Librarianship*, Vol. 46 No. 1, p. 102076.
- Calzada-Prado, F.J. and Marzal, M.A. (2013), “Incorporating data literacy into information literacy programs: core competencies and contents”, *Libri*, Vol. 62 No. 2.

- Carlson, J., Fosmire, M., Miller, C.C. and Megan Sapp Nelson, M.S. (2011), "Determining data information literacy needs: a study of students and research faculty. Libraries faculty and staff scholarship and research", available at: www.docs.lib.purdue.edu/lib_fsdocs/23 (accessed 22 May 2022).
- Chawinga, W.D. (2019), *Research Data Management in Public Universities in Malawi*, University of Western Cape, Bellville.
- Chawinga, W.D. and Zinn, S. (2019), "Global perspectives of research data sharing: a systematic literature review", *Library and Information Science Research*, Vol. 41 No. 2, pp. 109-122.
- Chiware, E.R. (2020), "Data librarianship in South African academic and research libraries: a survey", *Library Management*, Vol. 41 Nos 6/7, pp. 401-416.
- Chiware, E.R. and Becker, D.A. (2018), "Research data management services in Southern Africa: a readiness survey of academic and research libraries", *African Journal of Library, Archives and Information Science*, Vol. 28 No. 1, pp. 1-16.
- Elsayed, A.M. and Saleh, E.I. (2018), "Research data management and sharing among researchers in Arab universities: an exploratory study", *IFLA Journal*, Vol. 44 No. 4, pp. 281-299.
- Federer, L.M., Lu, Y.-L. and Joubert, D.J. (2016), "Data literacy training needs of biomedical researchers", *Journal of the Medical Library Association: JMLA*, Vol. 104 No. 1, p. 52.
- Fontichiaro, K. and Oehrli, J.A. (2016), "Why data literacy matters", *Knowledge Quest*, Vol. 44 No. 5, pp. 21-27.
- Gummer, E.S. and Mandinach, E.B. (2015), "Building a conceptual framework for data literacy", *Teachers College Record: The Voice of Scholarship in Education*, Vol. 117 No. 4, pp. 1-22.
- Kahn, M., Higgs, R., Davidson, J. and Jones, S. (2014), "Research data management in South Africa: how we shape up", *Australian Academic and Research Libraries*, Vol. 45 No. 4, pp. 296-308.
- Koltay, T. (2015), "Data literacy: in search of a name and identity", *Journal of Documentation*, Vol. 71 No. 2, pp. 401-415.
- Koltay, T. (2017), "Data literacy for researchers and data librarians", *Journal of Librarianship and Information Science*, Vol. 49 No. 1, pp. 3-14.
- Kubovics, M. and Zaušková, A. (2020), "Data literacy across target groups", in Kvetanová, Z. and Solík, M. and M. Z. Kubovics, A (Eds), *Megatrends and Media: On the Edge*, University of Ss. Cyril and Methodius, Trnava.
- Lyon, L., Ball, A., Duke, M. and Day, M. (2011), "Community capability model framework", available at: www.academia.edu/download/30836993/CCMDIRWhitePaper24042012.pdf (accessed 14 September 2022).
- Majid, S., Foo, S. and Zhang, X. (2018), "Research data management by academics and researchers: perceptions, knowledge and practices", *International Conference on Asian Digital Libraries*. Springer, Cham.
- Mandinach, E.B. and Gummer, E.S. (2016), "What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions", *Teaching and Teacher Education*, Vol. 60, pp. 366-376.
- Muronga, A. and Ogunlaja, A. (2022), *Data Literacy is as Important as Any Other Literacy*, University World News, Africa Edition, available at: www.universityworldnews.com/post.php?story=20220223090745688 (accessed 16 March 2022).
- Onyancha, O.B. (2018), "Navigating the rising metrics tide in the 21st century: which way for academic librarians in support of researchers in Sub-Saharan Africa?", available at: www.sajjils.journals.ac.za (accessed 13 May 2022).
- Onyancha, O.B. (2020), "Knowledge visualization and mapping of information literacy, 1975–2018", *IFLA Journal*, Vol. 46 No. 2, pp. 107-123.
- Patterson, L.H. (2016), "Research data management practices of emerging researchers at a South African research council", PhD Thesis, University of Pretoria.

- Patterton, L., Bothma, T.J. and Van Deventer, M.J. (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*, Vol. 84 No. 2, pp. 14-26.
- Prado, J.C. and Marzal, M.Á. (2013), "Incorporating data literacy into information literacy programs: core competencies and contents", *Libri*, Vol. 63 No. 2, pp. 123-134.
- Schneider, R. (2013), "Research data literacy", *Worldwide Commonalities and Challenges in Information Literacy Research and Practice: European Conference on Information Literacy, ECIL 2013 Istanbul, Turkey, October 22-25, 2013 Revised Selected Papers 1*. Springer International Publishing. pp. 134-140.
- Shied, M. (2004), "Information literacy, statistical literacy and data literacy", *IASSIST Quarterly Summer/Fall*.
- South African – European Union (2018), South African – European Union Dialogue Report. Department of Science and Technology, Pretoria.
- Stephenson, E. and Caravello, P.S. (2007), "Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project", *Reference Services Review*, Vol. 35 No. 4, pp. 525-540.
- United Nations Educational, Scientific and Cultural Organisation (UNESCO) International Bureau of Education (2022), "Multiple literacies", available at: www.ibe.unesco.org/en/glossary-curriculum-terminology/m/multiple-literacies (accessed 22 May 2022).
- United Nations Educational, Scientific and Cultural Organization (UNESCO) Institution for Statistics (2022), "Literacy: definition", available at: www.uis.unesco.org/en/glossary-term/literacy (accessed 22 May 2022).
- Universities South Africa (USAf) (2015), "Universities South Africa strategic framework 2015 – 2019", available at: www.usaf.ac.za/strategic-framework/ (accessed 13 April 2022).
- Vilar, P. and Zabukovec, V. (2018), "Research data management and research data literacy in Slovenian science", *Journal of Documentation*, Vol. 75 No. 1, pp. 24-43.

Appendix. ZA Digital literacy and Research Data Management Survey

You are invited to participate in a survey which aims to collect data about the data literacy of academics and research students in higher education institutions. From your responses we will be able to fully understand the current levels of awareness and gaps in knowledge which will help us develop appropriate data literacy training for the higher education community. Please answer all the questions and note that this survey is anonymous. It will take approximately 10 minutes to complete the entire survey. By completing this survey, you are consenting to the use of your data for research and dissemination purposes.

If you have any questions or comments as you are going through the survey, please contact the undersigned.

Thank you very much for your cooperation!

This survey is anonymous. The record of your survey responses does not contain any identifying information about you, unless a specific survey question explicitly asked for it. If you used an identifying token to access this survey, please rest assured that this token will not be stored together with your responses. It is managed in a separate database and will only be updated to indicate whether you did (or did not) complete this survey. There is no way of matching identification tokens with survey responses.

PART I: DEMOGRAPHIC INFORMATION

1. Your current primary role (choose one of the following answers)

Academic staff	
Research Student	
Other	

2. Your age (choose one of the following answers)

18 - 25	
26 - 35	
36 - 45	
46 - 55	
56 - 65	
65+	
Don't want to disclose	

3. Your discipline (please choose)

Natural Sciences - Mathematics	
Natural Sciences – Computer & Information Sciences	
Natural Sciences – Physical Sciences	
Natural Sciences – Chemical Sciences	
Natural Sciences – Earth and related Environmental Sciences	
Natural Sciences – Biological Sciences	
Engineering & Technology – Civil	
Engineering & Technology – Electrical	
Engineering, Electronic Engineering, Information Engineering	
Engineering & Technology – Mechanical	
Engineering & Technology – Chemical	
Engineering & Technology - Materials	
Engineering & Technology - Medical	
Engineering & Technology – Environmental	

(continued)

Data literacy training needs

Engineering & Technology – Environmental Biotechnology	
Engineering & Technology – Industrial	
Engineering & Technology – Nano-Technology	
Medical & Health Sciences – Basic Medicine	
Medical & Health Sciences – Clinical Medicine	
Medical & Health Sciences – Health Sciences	
Medical & Health Sciences – Health Biotechnology	
Agricultural Sciences – Agriculture, Forestry and Fisheries	
Agricultural Sciences – Animal and Dairy Science	
Agricultural Sciences – Veterinary Science	
Agricultural Sciences – Agricultural biotechnology	
Social Sciences - Psychology	
Social Sciences – Economics & Business	
Social Sciences – Educational Sciences	
Social Sciences – Sociology	
Social Sciences – Law	
Social Sciences – Political Science	
Social Sciences – Social and economic geography	
Social Sciences – Media & Communications	
Humanities – History & Archaeology	
Humanities – Languages & Literature	
Humanities – Philosophy, ethics and religion	
Humanities – Art (arts, history of arts, performing arts and music)	
Other	

4. Your legal gender

Please choose only one of the following:

Male	
Female	
Other	
Don't want to disclose	

5. Your country

Please write your answer here:

6. Your institution

Please write your answer here:

PART II: DATA LITERACY

7. Have you attended any formal data literacy training in the past?

No formal data training	
Some formal data training	

(continued)

8. Have you had a formal training on the following?
[Please choose all that apply]

Data Management Plan	
Metadata	
Consistent file naming	
Version control of data sets	
Data citation styles	

9. Is it necessary to have formal data literacy training?

Yes	
No	

10. Would you like to have a formal training on the following:
[Please choose all that apply]

Data Management Plan	
Metadata	
Consistent file naming	
Version control of data sets	
Data citation styles	

Thank you for your participation.

Corresponding author

Mathew Moyo can be contacted at: mathew.moyo@nwu.ac.za