IJDC | *General Article*

Mobilising a Nation: RDM Training and Education in South Africa

Refiloe Matlatse
University of Pretoria

Heila Pienaar
University of Pretoria

Martie van Deventer
CSIR

Abstract

The South African Network of Data and Information Curation Communities (NeDICC) was formed to promote the development and use of standards and best practices among South African data stewards and data librarians (NeDICC, 2015). The steering committee has members from various South African HEIs and research councils. As part of their service offerings NeDICC arranges seminars, workshops and conferences to promote awareness regarding digital curation. NeDICC has contributed to the increase in awareness, and growth of knowledge, on the subject of digital and data curation in South Africa (Kahn et al., 2014). NeDICC members are involved in the UP M.IT and Continued Professional Development training, and serve as external examiners for the UCT M.Phil in Digital Curation degree. NeDICC is responsible for the Research Data Management track at the annual e-Research conference in SA[1] and develops an annual training-focussed programme to provide workshop opportunities with both SA and foreign trainers. This paper specifically addresses the efforts by this community to mobilise and upskill South African librarians so that they would be willing and able to provide the necessary RDM services that would strengthen the national data effort.

1 eResearch conference: http://www.eresearch.ac.za/

Received 20 October 2016 ~ *Revision Received* 01 February 2017 ~ *Accepted* 01 February 2017

Correspondence should be addressed to Heila Pienaar, University of Pretoria, PO Box 002, Pretoria, 0001. Email: heila.pienaar@up.ac.za

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v12i1.579

Introduction

The need for Research Data Management (RDM) in South Africa (SA) intensified after the National Research Foundation (NRF) released a statement, which mandates that research data generated from publicly funded research projects must be deposited in a publicly accessible data repository with a digital object identifier (DOI) (NRF, 2015). The need for RDM is not unique to South Africa. Countries such as the United Kingdom (UK), the United States of America (USA), Australia and Denmark have long realised the need for RDM. Research funding bodies in these countries have also mandated that publicly funded research data be published in open data repositories for discovery and re-use by other researchers (Kahn et al., 2014).

In addition to the research funding bodies' mandates to make publicly funded research data available, there are other benefits and drivers for managing research data. Having research publicly available will aid in validating and authenticating research findings, research data will be re-used and will potentially improve the quality of research and improve research impact (Davidson, 2014). RDM will also facilitate interdisciplinary and/or collaborative research (Kahn et al., 2014).

Higher Education Institutions (HEIs) have shifted the responsibility of managing the research data to libraries and Library and Information Science (LIS) professionals. These professionals are considered the most skilled and/or experienced to support the management of research data in their institutions (CILIP, 2014).

Libraries already offer a full range of traditional library services such as library and information literacy training, developing library collections, cataloguing, as well as conducting literature searches for students, staff and researchers across various faculties and disciplines (CILIP, 2014). Libraries and librarians can extend their service offerings by collaborating with administrative services and support services across their institution, to provide RDM services. These can include playing an advocacy role within their institutions to develop a data policy, providing data literacy training to students and staff, offering reference services, and developing and maintaining data repositories to house data collections (Akers and Doty, 2012; Lyon, 2012).

Librarians have experience working with researchers and they work with information on a daily basis. Unfortunately, their past experience of working with the information and data is not enough. Further training and education is required to help LIS professionals, as well as any other professionals interested in, or required to support or facilitate, the management of research data.

Training and education can be presented in numerous forms, including formal university level undergraduate and/or postgraduate education, continued professional development (CPD) and informal face-to-face or online training.

The South African Network of Data and Information Curation Communities (NeDICC) was formed to promote the development and use of standards and best practices, to ensure the usability of digital material in support of e-Research among South African data stewards and data librarians (NeDICC, 2015). NeDICC develops an annual training-focussed programme to provide workshop opportunities with both SA and foreign trainers. As a result, the following workshops were presented during 2016: The life of research data and a roadmap to enable the implementation of services to support RDM; Data management planning; The role of the information professional in researcher engagement; Evaluation of research data repository applications; and a

library carpentry initiative focussing on the cleaning of data with Open Refine. Although a wider context is provided, this paper specifically addresses the efforts by this community to mobilize and upskill South African librarians so that they would be willing and able to provide the necessary RDM services that would strengthen the national data effort. There are, however, a number of formal training alternatives.

Formal Tertiary Data Science Training in South Africa

The Department of Science and Technology (DST) in South Africa, approved the establishment and funding of a National, e-Science, Postgraduate Teaching and Training Platform in September 2016. A curriculum for Big Data training will be made available for implementation in 2018.

Career entrants are currently able to pursue a career in, or related to, data management by enrolling for either an undergraduate or postgraduate degree. At undergraduate level, Sol Plaatjie University (SPU) offers a Bachelor of Science degree in Data Science. SPU was one of the first universities in SA to offer such a degree at undergraduate level. The three year programme is not specifically aimed at practicing LIS professionals, but rather at individuals who are interested in data science. The degree equips students with the competencies necessary to solve big data issues (SPU, 2017).

The Library and Information Study Centre (LISC) at the University of Cape Town (UCT) is presenting several courses in data curation and RDM. A short course in RDM is also offered annually. More importantly, it is the first university in Africa to offer a full Master's degree (MPhil) specialising in digital curation (Kahn et al., 2014; UCT, 2015). The programme is offered over two years, and students are required to complete coursework, as well as a mini-dissertation on a topic related to data management. Coursework includes: RDM, Information architecture and metadata, Technology platforms, and digital curation principles, theory and philosophy (UCT, 2015).

This degree targets data stewards – preferably new entrants, however practising librarians are also accepted into the programme.

The Information Science department at the University of Pretoria (UP) is responsible for a Carnegie-funded training programme for African Librarians. A Master's in Information Technology (M.IT) in Librarianship (Stream B) and a Continuing Professional Development (CPD) programme for Librarians, are offered. Both programmes include one module in RDM training. The university has also just introduced an M.IT (Stream C) in Big Data Science degree – as of February 2017 (University of Pretoria, 2017). This newly developed programme will be presented over two-three years, and includes one elective module dedicated to data curation. Similar to the UCT degree described above, students enrolled in this programme will need to complete a research-based mini-dissertation as well as course work (University of Pretoria, 2017). Here, the aim is to train data scientists, and although data stewards may pursue this avenue it appears quite unlikely at this stage.

The University of Witwatersrand provides a BSc Honours in the field of Big Data Analytics.

These formal RDM training opportunities were slower to develop than the informal or ad hoc initiatives. The formal educational programmes may appeal to individuals with a first degree, or early-career LIS professionals who seek to specialise in data science/management. LIS professionals, who seek to update their knowledge in RDM,

 doi:10.2218/ijdc.v12i1.579

may consider studying on their own or enrolling in online courses. Many academic and research libraries have developed their own RDM training, and have made the material openly available online. Some of these are discussed in the next section.

Informal Training

The Digital Curation Centre (DCC) and Joint Information Systems Committee (JISC) in the UK have designed and piloted many training programmes for LIS professionals. The following initiatives were randomly selected and reviewed in detail: RDMRose (University of Sheffield, 2015), RDMRose Lite, Digital Curation (DC) 101: How To Manage Research Data, SupportDM/TraD (University of East London, 2015), DIY RDM Training Kit for Librarians, New England Collaborative Data Management Curriculum (NECDMC) and Research Data MANTRA (University of Edinburgh).

The content of these initiatives were reviewed to gain appropriate background knowledge, and to establish a benchmark for our own initiatives. The topics most common in the RDM training programmes mentioned above are to provide an overview/introduction to RDM, data management planning, the role that the library and LIS is expected to play in RDM, and data sharing and re-use. Institutions may further customise the training programmes to suit their needs. Training will provide LIS professionals with some of the knowledge and skills required to support RDM in HEIs. LIS professionals must have an understanding of the research life-cycle, and good knowledge of research funder's mandates, data policy and governance, metadata standards copyright and other intellectual property laws (Charbonneau, 2013). The skills required to support RDM include: online data citation, data licensing, data management planning, data organisations, data storage and security (CILIP, 2014; Guy, 2013; Rice, 2014).

What follows is a brief report on an experiential learning workshop that was designed to provide information specialists and librarians exposure to RDM. Matlatse (2015) documented the full detail.

RDM Training: A South African Case Study

The day-long RDM workshop was presented to LIS professionals on August 11, 2014 as a combined initiative between NeDICC and the DCC. James Mullins of Purdue University, USA, was also available to present a brief case study to participants. The objective of the workshop was to provide novice attendees with a general overview of research data, as well as an introduction to the RDM subject.

The topics covered as part of the workshop were aligned to the analysis of the interventions mentioned above: An introduction to RDM, benefits and drivers of managing research data; The impact RDM has had on the profile of Purdue University Library; A CARDIO exercise; The state of RDM in South Africa; DIRISA (Data Intensive Research Initiative of South Africa) and its implications on RDM in SA; Useful RDM tools and services; A DMPOnline talk and demonstration; and Developing a data roadmap for your institution.

The training intervention was interactive, giving the attendees an opportunity to interact with the presenters and each other.

Research Design

An embedded qualitative research design was used, combining a quasi-experimental design (non-randomised control group pre-test-post-test design) and a case study research design. The intention was to establish whether such a workshop would increase the level of comfort for these staff members – to such an extent that they could be encouraged to become involved in the RDM drive. Data was collected using online questionnaires. Individuals who registered for the training were sent an email explaining the purpose of the research, and a link to the online questionnaire (Q1). 30 individuals completed the questionnaire, a response rate of 52%. The individuals were informed that there would be a second questionnaire that they would be asked to complete. As was expected the second questionnaire had a reduced response rate (36%).

The figure below depicts the sequence of the experiment. In the first block, the most significant results attained from the first questionnaire are shown. In the second block the contents of the intervention are displayed. In the third block, the most significant changes resulting from the intervention are shown.

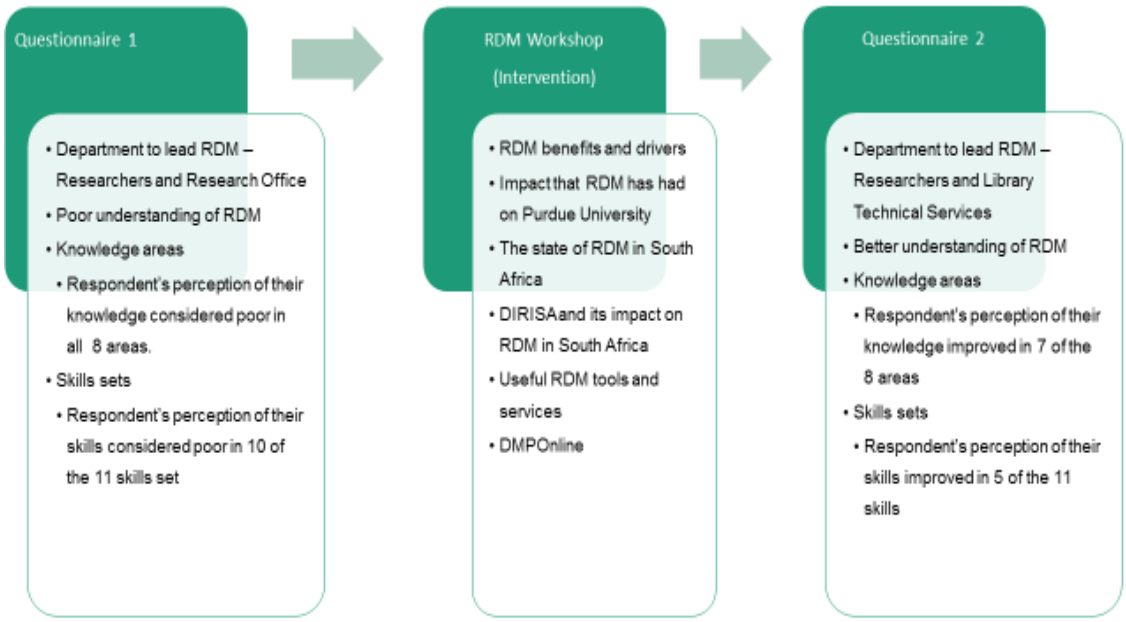

Figure 1. Research process followed with some results indicated.

The questionnaire themes were: demographic information, RDM policies at participants' institutions, RDM services at participants' institutions, roles and responsibilities for offering RDM services, understanding of RDM, disciplinary background necessary for RDM, knowledge and skills gaps, and feedback on the training intervention.

 doi:10.2218/ijdc.v12i1.579

Research Results

The majority of the participants were in the age range of 51 and above, and had over 21 years of experience. These would be the typical LIS professional who could be re-skilled to provide RDM support.

RDM Policies and Services at Participants' Institutions

At the time of the research, 53% of the respondent's reported that their institutions did not have any RDM policies, while 47% of the respondents reported that their institutions did offer RDM services.

The respondents were asked to list RDM services that they thought their institutions should offer. The respondents' answers were categorised as infrastructural services, consultative services and/or applied services.

The infrastructural services that the respondents thought their institutions should offer were: Providing data storage space; Providing IT hardware; Providing cloud services; and Developing a repository (for institutional data).

The consultative services the respondents thought their institutions should offer, were: Policy development; Data management planning; and offering advice on: file formats, Publishing, database design, data modelling, data sharing, data re-use rights, open access, and metadata standards.

The applied services the respondents thought their institutions should offer, were: Providing technical support; Data curation/actively managing data; Providing preservation services; Providing training; Creating awareness; Administering metadata to research data; Digitisation services; Designing data workflows and processes; Training; Creating guides/self-help manuals; Facilitating RDM; and creating awareness.

The respondents were also asked which departments they thought should lead the RDM initiatives in their institutions. They could select any three. The responses are tabled below.

Table 1. Shift in perception regarding the RDM 'owner'.

Department to lead RDM	Questionnaire 1	Questionnaire 2
IT Department	8 (27%)	11 (52%)
Library Technical Services	11 (37%)	14 (67%)
Publishers	3 (10%)	7 (33%)
Research Funding Bodies	8 (27%)	10 (48%)
Research Office	14 (47%)	13 (62%)
Researcher/Research Team	13 (43%)	14 (67%)
Subject librarian/Information Specialist	9 (30%)	13 (62%)
Other	1 (10%)	1 (5%)

When completing the first questionnaire, the respondents thought that the 'Research Office' (47%), the 'Researchers/Research Teams' (43%), and the 'Library Technical Services' (37%) should lead the RDM initiative. Thirty percent (30%) of the respondents thought that Subject Librarians/Information Specialists should lead the initiative. When completing the second questionnaire, after the training intervention, the 'Library Technical Services' (67%) and the 'Researchers/Research Team' (67%) became the most preferred departments to lead the RDM initiative. This was followed by the 'Research Office' and the 'Subject Librarian/Information Specialists' at 62% each. This led to the observation that the training intervention helped the respondents to better understand the role that LIS professionals can plan in supporting RDM. When looking at the responses in the second questionnaires, it was also possible to concluded that the respondents became aware of the roles that each of the departments could partake in, to manage research data.

RDM Understanding and Disciplinary Background of LIS Professionals

When answering the first questionnaire, 61% of the respondents perceived their RDM understanding to be poor at first observation. When completing the second questionnaire, the respondents' understanding of RDM had improved significantly however, 38% of the respondents still perceived their understanding of RDM to be poor.

The respondents were asked if they thought their background and past experience as LIS professionals would help them support or facilitate RDM at their institutions. At the first observation (questionnaire 1), 30% of the respondents thought their past LIS experience would assist. This improved slightly to 38% at the second observation (questionnaire 2). Most of the respondents indicated that this was due to a 'lack of formal RDM-related education' and the 'lack of practical experience'. The respondents also noted that the amount of information available on the topic of RDM is overwhelming, leading to information overload.

Gaps in RDM Knowledge

The respondents were asked in both questionnaires to rate their RDM (related) knowledge on a scale of one-four, one being poor and four being very good:

 doi:10.2218/ijdc.v12i1.579

Table 2. Shift in perceived knowledge regarding RDM.

RDM knowledge area	Poor knowledge		Good Knowledge		Difference in perception
	Q1 (%)	Q2 (%)	Q1 (%)	Q2 (%)	
Awareness of research data management	50	33	50	67	17
Long-term preservation	63	52	37	48	11
Policy and governance of data	77	71	23	29	6
Copyright and other IP rights	60	48	40	52	12
Metadata standards for data	57	52	43	48	5
Trusted repositories	53	62	47	38	-9
Funder mandates	90	67	10	33	23
Research process/lifecycle	50	19	50	81	31

The majority of the respondents perceived their RDM knowledge as poor (rated 1-2). After the training intervention, there was an improvement in the respondent's perception of their knowledge in seven out eight knowledge areas: Awareness of research data management; Long-term preservation; Policy and governance of data; Copyright and other IP rights; Metadata standards for data; Funder mandates; and Research process/lifecycle.

There was a significant reduction in the number of respondents who perceived their knowledge of 'Trusted Repositories' as poor.

Gaps in RDM Skills

We asked the respondents to rate their RDM skills on a scale of 1-4, with one being poor and four being very good. When answering the first questionnaire, the respondents perceived their RDM skills to be poor in all 11 areas.

The intervention had a mixed impact on the respondent's perceptions of their RDM skills. Unexpectedly, it had the opposite effect in RDM skills areas. Of the 11 RDM skills listed, there was an improvement in five skills, namely: Data Management Planning (DMP); Data licensing; Data appraisal and selection; Organisation and documentation of data; and managing data repositories.

However, there was a decline in respondents' perceptions of their RDM skills in five areas, namely: Indexing [adding metadata] skills; Online citation/referencing for data; Preparing data for deposit to repository; Data storage and security; Create guides and training materials for researchers.

The respondents' perception of their skills in 'Working with data repositories', remained unchanged.

Feedback (Second Questionnaire Only)

The respondents were asked to rate the usefulness of the RDM training on a scale of 1-4, one being 'not at all useful' and four being 'very useful'. 67% provided a rating of

four while 33% of the respondents provided a rating of three. The training intervention was therefore successful in improving the respondents' RDM understanding and knowledge. The respondents noted that they found that they had learnt more about the roles that libraries and librarians will play in terms of RDM services offerings. They also found it useful to learn about the various national and international RDM initiatives. The presentation by DIRISA was particularly useful in teaching the respondents about RDM within the broader South African initiative.

The training was interactive, which gave the attendees an opportunity to discuss their own institution's RDM-related initiatives. This gave way to an insightful discussion from an employee of the Cape Peninsula University of Technology and their particular RDM journey. The exercises that were part of the training helped the respondents to understand the importance of institutional planning.

The respondents were asked to describe what they would have liked to have been included as part of the training, but that was not addressed. They indicated that they would have liked more practical examples of RDM, and examples of both good and bad RDM practices. They would also have liked to see demonstrations of data management, and more detail regarding the topics of data citation, digital curation, and repositories.

As the training was presented by international experts, the respondents noted that handouts prior to the presentations would have helped them to follow the presentations better, especially because the speakers are perceived to have an accent, and there were some technical issues with sound. Some of the respondents found that there was too much content to focus on. They suggested extending the workshop over two days to allow the attendees enough time to understand the concepts. Another suggestion was that future training should be provided over an extended period of time, where training sessions were shorter in duration and focused only on individual topics or concepts at any one intervention. LIS professionals would therefore gain their knowledge over an extended period of time, however the same content would be disseminated.

Discussion

The research found that the RDM workshop was highly successful in enhancing the participant's perception of their RDM understanding and knowledge. The RDM workshop was less successful in enhancing the participant's perception of their RDM skills.

It was recommended that LIS professionals:

1. Take advantage of the online RDM training materials available, to enhance their understanding and knowledge of RDM;

2. Attend face-to-face training interventions to enhance, or develop, their RDM skills; and

3. Enrol in university level educational programmes to gain a qualification in RDM, if they qualify.

It was also recommended that institutions that provide RDM training should focus on specific aspects of RDM instead of offering a general overview. This is then what NeDICC will continue to focus on. The 2017 programme of training is closely aligned with the 23-Things initiative developed by the the Research Data Alliance Libraries for Research Data Interest Group (RDA, 2017).

 doi:10.2218/ijdc.v12i1.579

The literature consulted while conducting the research reported that LIS professionals had the necessary disciplinary background to support and facilitate RDM. The results of this research contradicted the finding above. LIS professionals who participated in the research, did not think that they had the disciplinary background required. They stated that they did not have enough practical experience or education. There was a slight difference in this opinion after the training intervention, where 38% of the LIS professionals thought that their disciplinary background would aid in RDM support. The LIS professionals who thought they had the disciplinary background to support RDM, stated that they experience working with data, and could translate those skills into RDM support.

The training intervention was successful in improving the LIS professional's perception of their understanding of RDM. The training intervention was also effective in enhancing the LIS professional's perception of their RDM knowledge. The LIS professional's improved their knowledge in seven of the eight areas. The training intervention was less successful in enhancing the RDM skills of the LIS professionals. The intervention did manage to provide the participants with an opportunity to reflect on their own skills, compared to the skills required for enabling, and supporting, RDM.

Conclusion and Recommendations

It is clear that both formal and informal RDM education and training are thriving in South Africa.

It is also clear that the international SKA project is the funding focus area for the DST. Data stewards would, therefore, in the short term at least, accept responsibility for their own training. LIS professionals interested in RDM have a number of options available to them. They can enrol in formal or informal training. This can be done in person (in a classroom) or online. It is up to individuals to determine their knowledge and skills gaps before selecting appropriate training. There is a lot of training material available online, which makes it possible for interested LIS professionals to 'self-teach' the necessary skills.

LIS professionals who qualify to study at a university may choose to enrol in an academic programme with a specific focus on RDM or a related subject. The library (line managers) may choose to offer their employees the opportunity to attend training, such as the RDM workshop used for this research, if they are expected to support RDM. Libraries can also develop in-house RDM training.

After the evaluation of the NeDICC & DCC workshop, NeDICC's training has become more practical, e.g. the Library carpentry workshop. Training and workshops organised by NeDICC are carried by the community both in terms of presenters and funding. As such, NeDICC should perhaps become a more formal/legal organisation to be officially acknowledged for its role – also in the training and development of data stewards. In essence though, educators and trainers for all types of research data management should find ways to collaborate, and acknowledge, the role that each party is playing in the mobilisation of our data stewardship effort.

References

Akers, K.G. & Doty, J. (2012). Differences among faculty ranks in views on research data management. *IASSIST Quarterly, 36*(2), pp 16-20. Retrieved from http://www.iassistdata.org/sites/default/files/iqvol36_2_doty.pdf

Charbonneau, D.H. (2013). Strategies for data management engagement. *Medical reference services quarterly, 32*(3): 365-374. doi:10.1080/02763869.2013.807089

Chartered Institute of Library and Information Professionals (CILIP). (2014). Research Data Management Briefing Paper. Retrieved from https://www.cilip.org.uk/sites/default/files/media/document/2017/research_data_management_briefing_july_2014_0.pdf

Davidson, J., Jones, S., Molloy, L. & Kejser, U.B. (2014). Emerging good practice in managing research data and research information within UK universities. *Procedia Computer Science*, 33: 215-222. Retrieved from http://forskningsbasen.deff.dk/Share.external?sp=Sb1d5f839-3331-4ec9-9741-2ec6fd3e4bee&sp=Ska

Guy, M. (2013). RDM training for librarians. DCC RDM Services case studies. Retrieved from http://www.dcc.ac.uk/resources/developing-rdm-services

Kahn, M., Higgs, R., Davidson, J. & Jones, S. (2014). Research data management in South Africa: How we shape up. *Australian Academic and Research Libraries*, 45(4): 296-308. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00048623.2014.951910

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation, 7*(1): 126-138. Retrieved from http://www.ijdc.net/index.php/jidc/article/view/210

Matlatse, R.L. (2015). An evaluation of a structured training event aimed at enhancing the research data management (RDM) knowledge and skills of Library and Information Science (LIS) professionals in South African Higher Education Institutions (HEIS). University of Pretoria, Unpublished Master of Information Technology (B) mini-dissertation. Retrieved from http://hdl.handle.net/2263/58393

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

Network of Data and Information Curation Communities (NeDICC). (2015). NeDICC: Its role and plans for 2016. Retrieved from https://nedicc.files.wordpress.com/2015/12/lucia-nedicc-citation-workshop-v02-2015-12-06.pdf

 doi:10.2218/ijdc.v12i1.579

Research Data Alliance. (2017). Libraries for research data interest group. Retrieved from https://www.rd-alliance.org/groups/libraries-research-data.html

Rice, R. (2014). Research data MANTRA: A labour of love. *Journal of eScience Librarianship, 3*(1): 44-51. Retrieved from http://dx.doi.org/10.7191/jeslib.2014.1056

Sol Platjes University. (2017). Bachelor of science degree in data science. Retrieved from http://www.spu.ac.za/index.php/programs-3/

University of Cape Town. (2015). UCT LISC offers short course in research data management. Retrieved from http://www.lisc.uct.ac.za/news/uct-lisc-offers-short-course-research-data-management

University of East London. (2015). TraD - Training for data management at UEL 2015. Retrieved from http://www.uel.ac.uk/trad/

University of Pretoria. (2017). MIT in big data science (Stream C). Retrieved from http://www.up.ac.za/en/school-of-information-technology/article/2324622/mit-in-big-data-science-stream-c

University of Sheffield. (2015). About RDMRose. Retrieved from http://www.sheffield.ac.uk/is/research/projects/rdmrose