Archiving South African digital research data: How ready are we?

AUTHORS:
Margaret M. Koopman[1]
Karin de Jager[2]

AFFILIATIONS:
[1]South African Environmental Observation Network – Fynbos Node, Cape Town, South Africa
[2]Library and Information Studies Centre, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Margaret Koopman

EMAIL:
margaret@saeon.ac.za

POSTAL ADDRESS:
SAEON – Fynbos Node, Private Bag X07, Claremont 7735, South Africa

DATES:
Received: 19 Aug. 2015
Revised: 08 Feb. 2016
Accepted: 13 Feb. 2016

KEYWORDS:
research data management; long-term ecological data; data repositories; data sharing; data preservation

HOW TO CITE:
Koopman MM, De Jager K. Archiving South African digital research data: How ready are we? S Afr J Sci. 2016;112(7/8), Art. #2015-0316, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150316

Digital data archiving and research data management have become increasingly important for institutions in South Africa, particularly after the announcement by the National Research Foundation, one of the principal South African academic research funders, recommending these actions for the research that they fund. A case study undertaken during the latter half of 2014, among the biological sciences researchers at a South African university, explored the state of data management and archiving at this institution and the readiness of researchers to engage with sharing their digital research data through repositories. It was found that while some researchers were already engaged with digital data archiving in repositories, neither researchers nor the university had implemented systematic research data management.

Introduction

A number of articles published in this journal are pertinent to the topic of digital data archiving[1,2], in particular the need for the preservation of long-term ecological data sets, which are crucial for understanding the management of the South African environment[3]. Research data have not traditionally had a home in university libraries or university archives, and have instead remained the responsibility of research units and researchers, or, in some cases, have been archived in special collections associated with a particular research unit and its specialised focus.[4] Data are the currency of research; but analogue and digital research data generated within academia have largely been an invisible resource utilised within the research unit and shared with a select group of trusted colleagues, and consequently their management is poorly understood. Digital data may have various states – raw data, which probably contain errors, require verification and, without metadata, only have meaning within a research discipline, and, at the other end of the spectrum, analysed data with metadata that can be downloaded from a repository and understood more broadly across disciplines. Each research discipline produces unique data, which require a range of specialised metadata languages and ontologies as well as subject-focused management and archiving solutions.[5]

The international focus on research data makes it important for South African researchers and policymakers to engage with the imperatives of ensuring that data are managed in a way that enables long-term security and accessibility. Data have commercial and intrinsic value, and in both cases it is important that they are archived for future use, particularly because re-collecting data is costly, in both time and money.[6]

The international context

The Advanced Research Projects Agency Network (ARPANET) was established in 1969 specifically to enable researchers to share data between laboratories in geographically distant locations.[7] ARPANET was the template upon which the Internet was subsequently built. The ubiquity of the Internet was the cornerstone of the open access initiative[8] which raised the question of universal access to research, particularly publicly funded research. There are, however, fundamental underlying factors that have led to the current preoccupation with research data archiving:

- Global climate change research has alerted governments and researchers to the value of long-term ecological studies.[9]
- Garnering funding has become an extremely competitive exercise and major funders want evidence that the research has not previously been undertaken, that the data collected will be preserved, and that the research will be open to scrutiny.[10]
- Providing underlying data is regarded as a way to prevent fraud in research, as the findings in the publications are expected to have robust scientific data underlying the research.[11]
- There is global awareness that digital records are in danger of being lost, or have already been lost because of inadequate management and preservation initiatives.[12]

Concern about the accessibility of digital data is universal and a plethora of published articles on the topic can be identified in the literature. Numerous case studies have been published which report on surveys conducted among the researchers who generate the data to establish the fate of research data.[13-15] In each case the findings were similar: lack of institutional support for research data management, lack of suitable data repositories to archive data for the long term and no incentives or mandates in place to encourage systematic data archiving, resulting in researchers keeping their data within the research unit.

Compounding this situation are attitudes towards sharing data. On the one hand, there are defensible reasons for data sharing:

- creating opportunities for further integrated research[16]
- contributing to global research initiatives, e.g. natural resource use decision-making[17]
- preventing expensive duplication of research[18]

- verifying research findings[19]
- sharing data to make research more efficient[20] and to ensure continuation of research
- making research transparent[21]
- improving researchers' international profiles[22]

On the other hand, there are the cautious and often negative attitudes of the researchers who produce the data and who are slow to archive or make their data available.[23] Ecological researchers do not have a tradition of sharing research data, other than with trusted colleagues and collaborators. In his interview for a Data Matters blog from *Scientific Data*, Gavin Simpson, a Canadian environmental scientist, succinctly presented the point of view of ecologists: 'If you've toiled in the field for years to collect data then you're not going to be very easily convinced to make the data available. It's not part of our culture'.[24] It would appear that the only way to resolve the concerns around archiving and sharing research data in a formal repository is to make data archiving mandatory, to formalise data management and to ensure that data generators benefit from sharing their data. Digital Object Identifiers (DOIs) for data enable data users to acknowledge data generators in the same way that the authors of articles and books are acknowledged. Ensuring that data are available for long-term reuse, and that they can be acknowledged through DOIs will enable data generators to use data citations, in addition to article citations, when preparing funding proposals for further research.

A number of mainstream academic journals have made data archiving mandatory – *American Naturalist*, *Molecular Ecology*, *Nature*, the Public Library of Science (PLoS) journals, Royal Society of London journals and *Science*, to name a few in the field of ecology. Funder mandates are seen as the most reliable method for making data management and archiving a part of the research data life cycle[25] – the process whereby a researcher plans and documents the various steps in data creation, processing, and analysing as part of the research design. A data management plan includes the preservation of the data and a process whereby data can be shared and reused along with the detailed data description, or metadata, that must be archived with the data. A recent editorial in *Nature*[26] pertinent to open access publishing reveals that the Research Council UK, with oversight of seven public funders in the UK, has found that mandatory open access publishing continues to be problematic, with considerably less than 100% compliance. It is not surprising that archiving research data to make them openly accessible is in a far less developed state.

International initiatives that stand out in their response to digital data archiving initiatives include:

- The Digital Curation Centre (DCC) at Edinburgh University, established in 2004[27]
- The Data Archiving and Networked Services (DANS) in the Netherlands, established in 2004/2005[28]
- The Long-Term Ecological Research Network of the National Science Foundation in the USA, established in 1982[29]
- GenBank, the genetic sequence database provided by the US National Center for Biotechnology Information, established in 1982[30]

Numerous international solutions can be used by researchers to archive their data and by policymakers and institutional managers as examples of best practice for a range of research disciplines.[31] The growth in digital data repositories has resulted in the establishment of an international, peer-reviewed process – 'The Data Seal of Approval' – initiated at DANS, which enables institutions to evaluate the reliability of their repository.[32] A repository carrying the Data Seal of Approval is immediately recognisable to researchers and policymakers as a reliable source of data and a reliable site on which to deposit data.

The South African context

Several initiatives for archiving ecological data have been in operation in South Africa, such as the Southern African Data Centre for Oceanography (SADCO) that has been in existence since the 1960s; AfrOBIS, the African component of the international Ocean Biogeographic Information System (OBIS) that was set up in 1997 as a project of the International Oceanographic Commission; and the South African Bird Ringing Unit (SAFRING) which has been contributing to knowledge about bird migration since 1948. The main digital data archiving platform focusing on environmental data is that of the South African Environmental Observation Network (SAEON), established in 2002.

A survey was undertaken to investigate the state of data management and archiving within the Department of Biological Sciences at the University of Cape Town (UCT) and the readiness of researchers to engage with sharing their digital research data in repositories. It will be seen from the results of the survey reported below that these repositories are among those utilised by the academic researchers who were surveyed.

Survey of data archiving expertise and initiatives

Researchers from the Department of Biological Sciences at UCT participated in an online multiple-choice survey, designed to be both interrogative and informative, about their data management and archiving initiatives. The survey was a variation of the computerised self-administered questionnaire[33] – an anonymous web-based survey in which the respondents linked to an identified site and completed the questionnaire online without assistance. The survey was designed using Google Forms and consisted of 32 multiple-choice questions. The research was undertaken after ethical clearance from UCT (reference number UCTLIS201408-01).

Face-to-face interviews were conducted with a small group of research technicians and emeritus/retired researchers using the questions from the self-administered survey.

Out of an estimated target population of 318, a total of 163 researchers completed the survey. The survey was conducted over a 5-week period with weekly email reminders sent out to the target population.

To enable an understanding of the Department's researchers' data management issues and activities, the questions were divided into different categories:

- researcher characteristics
- researcher funding streams
- publishing characteristics
- data characteristics
- data ownership, intellectual property and copyright
- housekeeping routines and responsibilities
- long-term data potential, archiving and metadata
- institutional engagement and data management education possibilities

Researcher characteristics

The respondents were divided into seven categories of researchers: Emeritus/Retired (10 individuals); Academic (24 individuals); Research Associate (14 individuals); Postdoctoral (21 individuals); PhD (39 individuals); Master's (32 individuals); and Honours (9 individuals). An additional category of respondents – Other/Technical – consisted of 14 individuals made up of research technicians and research support staff.

The respondents were highly qualified, with 71% having either a master's degree or a doctorate. Among the Other/Technical category there were a number of PhD and master's graduates.

Researcher funding streams

Biological research is generally expensive to fund, particularly marine and Antarctic research that require ocean-going vessels that are not available through the university. Such research requires international collaboration and involvement in government initiatives which are publicly funded programmes. The proportion of public funding of the respondents' research is high: 73% of research is at least partially funded through public funds (Table 1). Such funding renders researchers accountable to the public to make their research openly available and to ensure that their data are available for future research.

Table 1: Percentage of public funding of respondents' research

Researcher category	Percentage of public funding					
	100%	75%	50%	25%	0%	Don't know
Emeritus/ Retired	5	1	3	1	0	0
Academic	14	5	0	1	4	0
Research Associate	4	4	0	2	3	1
Postdoctoral	15	0	3	0	1	2
PhD	18	6	6	0	5	4
MSc	12	0	3	1	7	9
Honours	3	0	1	2	2	1
Total	**71**	**16**	**16**	**7**	**22**	**17**

Information on co-funding through international collaboration was extracted by an examination of published research output during 2007, 2010 and 2014. This examination demonstrated that private and overseas co-funding matched public funding in 2007, and exceeded public funding in 2010 and 2014. The authors collaborating on one paper may have been co-funded by more than one party, resulting in more co-funding categories than total articles (Figure 1).

Publishing characteristics

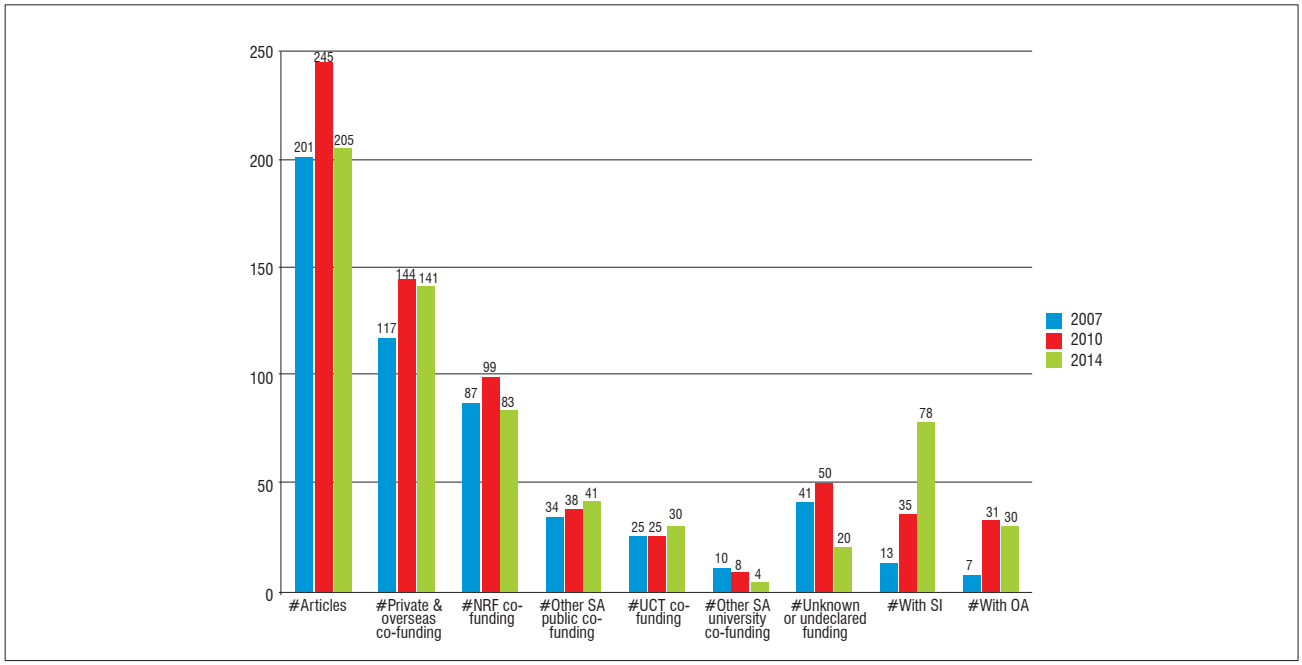
Academic research findings were made available in the past through the publications of learned societies. When learned societies ceased their publications, the task of publishing findings was taken on by discipline-specific journals published by commercial or not-for-profit scholarly publishers. In both cases research was largely hidden from the general public.

The trend for researchers to make their publications openly accessible has grown because of funding and collaboration mandates and/or to ensure that the research receives the widest audience possible. During 2014, open-access articles amounted to 15% of the total published output of the Department of Biological Sciences (Figure 1). In order to comply with future public funding mandates, the percentage of articles – with accompanying underlying data – would be expected to at least match the percentage of public funding.

An investigation was undertaken into the publication output in scientific journals of researchers in the Department of Biological Sciences, in parallel with the survey, to establish how many articles were published with supplementary information – for example data, code, images or extended bibliographies – as a way of sharing other products relating to the published research. In 2007, only 6% of the published papers were accompanied by supplementary information. This percentage climbed to 14% in 2010 and jumped to 38% in 2014 (Figure 1).

Data characteristics

It was found that researchers in the Department of Biological Sciences have worked with a range of long-term data sets, ranging from over 10 years to over 50 years in extent. Past research has generated digital data in many different formats, which have been archived on various media such as zip drives and 8-, 5¼- or 3½-inch floppy disks. Many digital data sets were in proprietary formats such as Lotus, dBase, Quattro Pro and other Corel products, or early versions of Microsoft, creating problems for long-term data accessibility. Emeritus, retired and senior academics reported data lost because of incompatibility with contemporary computer hardware, operating systems or software programs.



NRF, National Research Foundation; SI, supplementary information; OA, open access

Figure 1: Number of articles by researchers in the Department of Biological Sciences funded through private and overseas funding compared to those funded through public funding.

Other researchers had retained field or experimental notebooks which contained their raw data in analogue format. The majority (91%) of digital data formats generated by younger researchers were spreadsheets (in XML or CSV format). Improved management of digital data would prevent obsolescence and loss of data from this cohort of researchers.

Data were reused and shared within a controlled group of collaborating researchers. Very few researchers allowed open use of the data sets under their control, with the exception of the researchers in the Animal Demography Unit of the Department of Biological Sciences who managed large sets of 'citizen science' data that carried an open data mandate. Analogue data sets also existed within the department, but were largely invisible through lack of description and archiving. Many past students' data sets accompanied dissertations and remained as appendices in analogue theses, also lacking indexing and description. A number of retired research staff reported that their data either had been thrown away ($n=3$) or had nowhere to go ($n=5$) because of a lack of interest among colleagues and the institution. Early digital data had also been lost through lack of institutional support, foresight and responsible management. Instances of old digital data still in existence but inaccessible on contemporary computer platforms were common among senior academics. At the time of the investigation, there were no institutional plans in place to rescue these digital data sets.

Data ownership, intellectual property and copyright

Data ownership was found to be a key inhibitor to data sharing. Opinion about ownership varied: does the funder, the institution, the research unit, the supervisor, or the student own the data? Or is the owner a combination of all these potential data owners?

Researchers' responses varied according to the category in which they were placed, with 'Researcher ownership' scoring the highest by 'Emeritus/Retired', 'Academic' and 'Postdoctoral' respondents; 'UCT ownership' was scored highest by the 'PhD' category; and 'Supervisor ownership' of data was scored highest by the 'MSc' category. The highest score allocated by both 'Research Associate' and 'Honours' categories was for the response 'Don't know'. Other data owners were reported as 'the organisation I work for', 'citizen science observers', 'South African government', and 'open data from a repository'. Overall responses to data ownership can be seen in Figure 2.

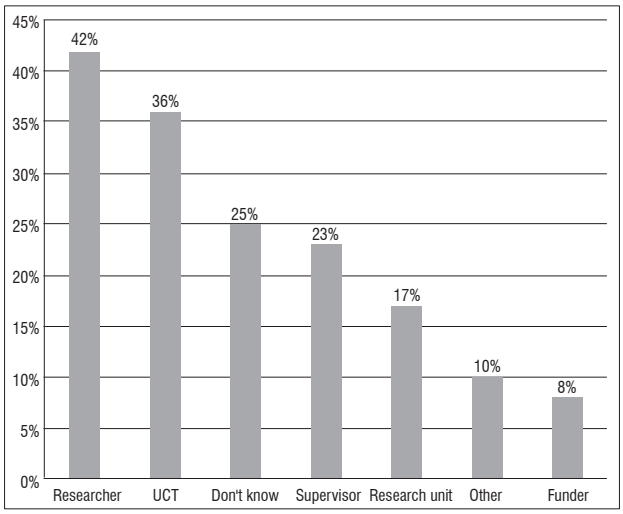


Figure 2: Percentage of respondents who attributed data ownership to each possible response.

When asked if their data should be made available for future research, 88% of researchers responded positively. But there were caveats to this response, which are reported in Table 2. The raw data revealed cascading requirements for making data available. For example, respondents were willing to share data only after publication and only if the data generator was offered co-authorship, or, after publication and only on request so that the data generator could evaluate the researcher and project wishing to use the data. Being able to trust the person with whom data would be shared was an important consideration.

Table 2: Conditions for sharing research data as reported by respondents

Condition for sharing data	% Respondents
Only after publication	62
Only on request so that I can discriminate	37
Only if open access with acknowledgement	30
Only if I am offered co-authorship	28
Only to a trusted researcher	25
Only if my data sets have DOIs	7
Only if data sets have a Creative Commons licence	3

The raw data demonstrated that the respondents who were prepared to share data through acknowledgement, inclusion of DOIs or through publishing under Creative Commons licences were those contributing to or utilising data sets such as 'citizen science' data that already had an open mandate.

In some cases, researchers indicated that there were copyright restrictions on the data they were using and that they were not permitted to share these data sets. In another case, a research group reported that their data had been misappropriated by another research group on campus, because no memorandum of understanding had been in place to specify agreed terms of data use.

Researchers were sharing data, but in the majority of cases this sharing was not through data repositories. Respondents to the survey could select multiple answers and reported sharing of data through the following methods:

- by email on request, 70%
- within published papers, 38%
- in the public domain, 17%
- through a collaborative initiative, 15%
- through a repository, 12%
- through the research unit's server, 3%

Until data ownership is resolved, through funding or collaborative agreements – and incentives such as the acknowledgement of data generators through the use of data DOIs are commonplace – data sharing will remain a contested issue.

Housekeeping routines and responsibilities

The survey interrogated data management and preservation activities such as storage, back-up routines and data migration routines. Questions on responsibility for data storage revealed a range of perceptions (Table 3) that focused data responsibility on the research unit rather than institutional IT departments, the university library or repositories. At the time of the survey, the institution took no responsibility for research data, although researchers could avail themselves of storage space on an IT server at a cost. Departmental IT personnel interviewed for the investigation (included in the Other/Technical category) indicated that they would give advice to researchers for the storage of data but that they were not responsible for researchers' data. Researchers and research units took responsibility for their data, as this was considered to be the status quo.

Table 3: The number of respondents who attributed responsibility for storage of data sets to each possible response

Data storage location	Number of respondents
Researcher/supervisor	95
Research unit	68
University library	37
National repository	33
Departmental IT personnel	28
International repository	18
Don't know	13
University IT department	12
Other	3

Researchers were diligent about back-up routines. The type and location of data back-ups is of interest as it demonstrates changing trends in data storage (Table 4). CD/DVDs are falling out of fashion and cloud storage is becoming more popular, although some researchers expressed reservations about data privacy on cloud storage. There appeared to be no consensus at the time of the survey, although external hard drives were the favoured medium, and keeping a back-up at home was the favoured location. One cannot predict how research data will be stored and backed up in the future. But, the move to cloud storage as a more accessible format which does not require the researcher to purchase or carry around additional hardware appears to be gaining popularity. Institutional commitment to research data management through the provision of staging repositories for active research data could improve the security of research data.

Table 4: Percentage of respondents who back up data at each location

Back-up location	% Respondents
Hard drive	83
PC/laptop	55
At home	42
Cloud storage	39
Office	26
Flash drive	25
Server	15
CD/DVD	4
Other	4
UCT ICTS	0.6

Long-term data potential, archiving and metadata

A range of data repositories was utilised by the researchers who were obliged to archive their data, either through collaborations, funding or publisher mandates, or through disciplinary mandates, such as for genome data. Repositories were also utilised by researchers to access data to use in their research. The repositories reportedly utilised by respondents are listed in Table 5.

Table 5: Repositories utilised by researchers in the Department of Biological Sciences

Repository	% Respondents
GenBank	23
SAEON	17
UCT Libraries Digital Repository	9
GBIF/SANBIF	8
Other†	8
EMBL	4
Dryad	3
JStor Global Plants	3
Movebank	3
AfrObis/Obis	2
SADCO	2

†Other repositories mentioned were a suite of Animal Demography Unit (UCT) databases, SANBI data archives, The British Library Sound Archive, and those of The National Marine Linefish System, The BirdLife Seabird Tracking Database, UvA-BiTS (University of Amsterdam Bird Tracking System) and Iziko Museum.

It was found that only 12% of the respondents had used a repository as a means of sharing data, although responses shown in Table 3 indicated that a higher percentage considered a repository to be the appropriate place for responsible data archiving. This apparent contradiction is understandable as routine data archiving in repositories was unknown to many researchers.

Table 6: Types of metadata considered important to describe research data as reported by respondents

Metadata field	Number of respondents
Name of creator/ research unit's name	91
Title of the data set	90
Geographical coordinates	87
Description of the data set	86
Contact details of creator/ research unit	78
Date of data creation	78
Collection methods	78
Taxonomic names	68
Beginning and end dates of project	57
Equipment used to gather data	57
Data format/s	54
Keywords	45
Names of funders	38
Copyright provisions	37
Title of umbrella project	31
Contact details of umbrella project	14
Don't assign metadata	14
Contact details of funders	9
Don't know	7

As metadata are a fundamental component of data sharing, the survey was devised to include a question for which there was a range of mandatory metadata fields as possible answers; respondents could select multiple answers. The possible answers and percentage of respondents who gave each answer are shown in Table 6.

The question that elicited the responses in Table 6 was also intended to sensitise researchers to metadata fields that could be used to describe their research, as the assignment of metadata was a new concept for many researchers at the time. Maintaining detailed descriptions about their data through the use of metadata did not appear to be a routine activity and a number of researchers indicated that they did not assign metadata. The fields shown in Table 6 represent those required by the Ecological Metadata Language (EML) standard.

Researchers were asked what they thought was the purpose of data curation (Table 7). Migrating data into formats that could be used by current software and operating systems received the lowest response. This was a neglected aspect of data management among senior academics that had resulted in data obsolescence instead of data remaining viable for long-term research.

Table 7: Purposes of data curation as reported by respondents

Purpose of data curation	% Respondents
Storing data for access and use	83
Ensuring that data are secure and backed up and available	79
Making sure data are available for future use	77
Maintaining research data in the long term to enable reuse	68
Ensuring that data are organised and indexed	54
Migrating data to new platforms/software	35

In order to build up long-term data sets for long-term ecological research such as land-use or climate change, data management will need to become an integrated part of the research life cycle.

Institutional engagement and data management education possibilities

The survey contained three questions posed in order to gauge the appetite of researchers for data management education. The questions and percentage responses can be seen in Table 8.

Table 8: Potential for data management education as indicated by researchers in the Department of Biological Sciences

Question	Answer		
	Yes	No	Other
Would you attend a workshop to discuss metadata generation?	50%	40%	Metadata are not applicable (10%)
Would you attend a workshop to discuss data management?	58%	28%	Would prefer an online resource (14%)
Do you require data management assistance?	51%	45%	Hire students to assist with data management (4%)

Various aspects of data management, such as information about metadata languages and standards, are an opportunity for librarians to develop online resources in support of data-generating disciplines. Tools such as LibGuides[34] are ideally suited for providing such online support for researchers. Although some data generators were using students to assist in data management, 51% of the respondents said that they 'would like more information about managing data efficiently'.

For UCT libraries to give appropriate support to researchers, librarians with specialised backgrounds or experience and the ability to interact with researchers would be required. Much of the advice needed at undergraduate and postgraduate levels is generic, such as file naming conventions, data back-up habits, keeping records of the what, how, when, where and why data were gathered (metadata), and types of metadata protocols required for archiving specific data types. The UCT libraries also have a role to play in directing researchers to other divisions on campus where information on topics – such as research funding, ethics support, IP support and temporary data storage – can be found.

During 2014 an eResearch Centre was established at UCT[35] which initiated a number of activities such as workshops and conferences to support research data generators. Collaborators in this initiative were UCT libraries, UCT ICTS (Information and Communication Technology Services) and the UCT Research Office who were in the process of developing a research data management policy for the university. The UCT libraries established a web presence for research data management (RDM) to advertise 'resources and training on research data management'[36].

Institutional managers have a role to play in ensuring that data management and curation are accounted for in research budgets. Whereas 88% of respondents indicated that their research should be made available for future research, only 18% budgeted for data management and data curation and only 26% had a data preservation plan.

Conclusion

The survey demonstrated that, even within the Department of Biological Sciences, research was varied and data collection and interpretation required a range of specialist skills, equipment and tools. Any discussion of metadata should include the standards and metadata languages appropriate for all types of research data in order for researchers to successfully describe their data for long-term preservation. The link between metadata and sharing has to be made in order for researchers to see the importance of comprehensive data descriptions, as without metadata, their data have no long-term value.

There had been no systematic interventions at UCT for supporting researchers with data management or data storage facilities, and an ad-hoc situation with varying success in the preservation of research data had been the status quo. Research data archiving for long-term preservation requires secure funding streams as well as training in RDM. Assistance with the development of RDM plans, soon to be required by South African research funders, is one of the ways in which the institution can assist researchers to apportion funding for data preservation.

Systematic RDM and archiving will only come about when proposed policies have been established in consultation with researchers. RDM education of the new cohort of researchers is a prerequisite for establishing systematic data archiving, and initiatives should be introduced at senior undergraduate level. Because RDM is a relatively new concept to South African researchers, support should also be offered to senior- and mid-level academic researchers so that they are sufficiently informed to ensure that student data are properly managed and archived. New research projects should include a data archiving and sharing plan as part of the overall project plan.

At the time of the survey, there was no strategy in place for the management or archiving of pre-digital or early digital research data, and some of these data were still in the hands of the retired and emeritus staff of the Department of Biological Sciences who were interviewed. Ensuring that long-term data sets are preserved is urgent and important, as it is not possible to recreate long-term ecological data because the impacts of human population expansion and resource usage change ecological systems over time.

Acknowledgements

We thank the Percy FitzPatrick Institute of African Ornithology for co-funding; the survey respondents for their participation; and the anonymous reviewers for their input in improving the manuscript.

Authors' contributions

M.M.K. was the project leader; M.M.K. and K.d.J. were responsible for project design; M.M.K. undertook the survey; K.d.J. made conceptual contributions and supervised the dissertation on which the article is based; M.M.K. prepared, analysed and documented the results of the survey; and M.M.K. and K.d.J. wrote the manuscript.

References

1. Grundlingh ML, Von St Ange UB, Bolton JJ, Bursey M, Compagno L, Cooper R, et al. AfrOBIS: A marine biogeographic information system for sub-saharan Africa. S Afr J Sci. 2007;103:91–93.
2. Van Jaarsveld AS, Pauw JC, Mundree S, Mecenero S, Coetzee BWT, Alard GF. South African Environmental Observation Network: Vision, design and status. S Afr J Sci. 2007;103:289–294.
3. Henschel J, Pauw J, Banyikwa F, Brito R, Chabwela H, Palmer T, et al. Developing the Environmental Long-Term Observatories Network of Southern Africa (ELTOSA). S Afr J Sci. 2003;99:100–108.
4. South African Bird Ringing Unit [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://safring.adu.org.za/content.php?id=1
5. Digital Curation Centre. Disciplinary metadata [homepage on the Internet]. No date [cited 2016 Jun 21] Available from: http://www.dcc.ac.uk/resources/metadata-standards
6. Kenall A, Edmonds S, Goodman L, Bal L, Flintoft L, Shanahan DR, et al. Better reporting for better research: A checklist for reproducibility. BMC Neurosci. 2015;16, Art. #44, 3 pages. http://dx.doi.org/10.1186/s12868-015-0177-z
7. Dasgupta S. Arpanet. In: Encyclopedia of virtual communities and technologies. London: Idea Group Reference; 2006. p. 173. http://dx.doi.org/10.4018/978-1-59140-563-4
8. Budapest Open Access Initiative [homepage on the Internet]. c2002 [cited 2016 Jun 21]. Available from: http://www.soros.org/openaccess/read.shtml
9. Walther GR, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, et al. Ecological responses to recent climate change. Nature. 2002;416:389–395. http://dx.doi.org/10.1038/416389a
10. The Editors. Editorial: Dr No Money: The broken science funding system. Sci Am. 2011 April 19; Mind. Available from: http://www.scientificamerican.com/article/dr-no-money/
11. Doorn P, Dillo I, Van Horik R. Lies, damned lies and research data: Can data sharing prevent data fraud? Int J Digit Curation. 2013;8:229–243. http://dx.doi.org/10.2218/ijdc.v8i1.256
12. Computer History Museum. Digital Dark Age: Revolution preview [video on the Internet]. c2011 [cited 2016 Jun 21]. Available from: https://www.youtube.com/watch?v=PSlMzirvsFc
13. Diekmann F. Data practices of agricultural scientists: Results from an exploratory study. J Agr Food Inform. 2012;13:14–34. http://dx.doi.org/10.1080/10496505.2012.636005
14. Elliot G. Otago biodiversity data management project report. Part 1. Questionnaire report. Dunedin: University of Otago Library; 2008. Available from: http://otago.ourarchive.ac.nz/handle/10523/198
15. Scaramozzino JM, Ramírez ML, McGaughey KJ. A study of faculty data curation behaviors and attitudes at a teaching-centered university. Coll Res Libr. 2012;75:349–365. http://dx.doi.org/10.5860/crl-255
16. Costello ML. Motivating online publication of data. BioScience. 2009;59:418–427. http://dx.doi.org/10.1525/bio.2009.59.5.9
17. Huang X, Hawkins BA, Qiao G. Biodiversity data sharing: Will peer-reviewed data papers work? BioScience. 2013;63(1):5–6. http://dx.doi.org/10.1525/bio.2013.63.1.2
18. Fry J, Lockyer S, Oppenheim C. Identifying benefits arising from the duration and open sharing of research data produced by UK higher education and research institutes [programme/project deposit]. c2009 [cited 2016 Jun 21]. Available from: http://repository.jisc.ac.uk/279/
19. Borgman CL. The conundrum of sharing research data. J Am Soc Inform Sci Technol. 2012;63:1059–1078. http://dx.doi.org/10.1002/asi.22634
20. Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS ONE. 2011;6(7), Art. #e18657, 13 pages. http://dx.plos.org/10.1371/journal.pone.0018657
21. Molloy JC. The Open Knowledge Foundation: Open data means better science. PLoS Biol. 2011;9(12), Art. #e1001195, 4 pages. http://dx.plos.org/10.1371/journal.pbio.1001195
22. Piwowar HA, Day RS, Fridsma DB, Ionnidis J. Sharing detailed research data is associated with increased citation rate. PLoS ONE. 2007;2(3), Art. #e308, 5 pages. http://dx.plos.org/10.1371/journal.pone.0000308.
23. Van Noorden R. Confusion over open-data rules. Nature. 2014;515:478. http://dx.doi.org/10.1038/515478a
24. Hufton A. Data matters: Interview with Gavin Simpson. Scientific Data Updates. 2014 August 26. Available from: http://blogs.nature.com/scientificdata/2014/08/26/data-matters-interview-with- gavin-simpson/
25. Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, et al. Mandated data archiving greatly improves access to research data. FASEB J. 2013;27:1304–1308. http://dx.doi.org/10.1096/fj.12-218164
26. Editorial: All that glitters. Nature. 2015;520:131. http://dx.doi.org/10.1038/520131a
27. Digital Curation Centre [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.dcc.ac.uk/
28. Data Archiving and Networked Services (DANS) [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.dans.knaw.nl/en
29. The Longterm Ecological Research Network (LTER) [homepage on the Internet]. c2013 [cited 2016 Jun 21]. Available from: http://www.lternet.edu
30. GenBank overview [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.ncbi.nlm.nih.gov/genbank/
31. Registry of Research Data Repositories [homepage on the Internet]. c2012 [cited 2016 Jun 21]. Available from: http://www.re3data.org/
32. Data Seal of Approval [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://datasealofapproval.org/en/information/about/
33. Babbie E, Mouton J. The practice of social research. Cape Town: Oxford University Press; 2001.
34. SpringShare. LibGuides [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://springshare.com/libguides/
35. University of Cape Town. eResearch Centre [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.eresearch.uct.ac.za
36. University of Cape Town Libraries. Research data management [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.digitalservices.lib.uct.ac.za/dls/rdm