UCT Research Data Management Policy Project: Report

Lynn Woolfrey
March 2014

CONTENTS

1 BACKGROUND

The University of Cape Town Research Committee (URC) initiated the process of institution-wide data curation planning at the University of Cape Town (UCT). This arose from URC meeting

discussions concerning research support at UCT. The URC wanted to follow international best practice and develop policy to archive and manage the data collected by UCT researchers in the process of their research. Creation and implementation of Research Data Management (RDM) policies at universities has become a prerequisite for research integrity and research efficiency at these institutions. URC members undertook research in this area to inform policymaking. Sakkie Janse van Rensburg, ICTS Director, and Prof Ed Rybicki of the URC visited Australia to investigate research data storage at Australian universities and brought back information on best practice at these institutions. The UCT Director of Libraries, Gwenda Thomas, scoped data management policies at South African universities (University Research Committee, 2012). This revealed that only UP had an official Research Data Management policy, which was sparse on detail.

The URC put together a Task Team to help UCT establish an effective Research Data Management (RDM) policy. The Task Team included UCT's ICT Director and the Director of Libraries, as well as the Deputy Deans for Research in each faculty, or their nominees. (University Research Committee, 2013). The Research Data Management Task Team (RDMTT) also co-opted a representative from UCT's Research Contracts and Intellectual Property Services (RCIPS) and the Director and Manager of UCT's data service, DataFirst. DataFirst's Manager agreed to lead the project, and together with the DOL, drew up a Project Plan (PP).

The Project Plan was presented to the URC by the Project Coordinator at a meeting in April 2013 Because of the policy would have cost implications for the university, the meeting suggested the project plan be submitted to UCT's Senate Executive Committee (SEC) for approval, before the work was taken further. The SEC noted that the URC initiative was a necessary response to requirements by funders that research data be kept and be accessible. Issues addressed by the SEC were the repository needs of such a data collection project; data protection; and data sharing. Cost implications were raised, with the URC's suggested cost of 20 million. These costs related to storage infrastructure, as well as data collation and metadata creation costs. The SEC provided in-principle support for the project, which was to include a draft policy document and implementation plan.

2 RESEARCH DATA MANAGEMENT STAKEHOLDER WORKSHOP

June 2013

Policymaking required broad based consultation with stakeholders in RDM at the university. In this regard the project milestones included the tabling of RDM policy documentation at various stages at workshops and meetings. A Research Data Management workshop was held on 20130624 to obtain feedback from key stakeholders. The report of the workshop is provided as APPENDIX 1 of this document.

3 SCOPING STUDY OF RESEARCH DATA MANAGEMENT POLICIES OF FUNDING AGENCIES

June 2013

This scoping study was undertaken by Warda Sablay of Research Contracts and Intellectual Property Services (RCIPS) at UCT. Ms Sablay presented the results of her study at the Workshop in June. This served to isolate common themes and highlight best practice. Funders' policies are seen to carry weight with researchers, and offer a means of incentive for researcher support for RDM, so this information was an important input to policy creation. The workshop presentation was an overview of

RDM requirements of key funders of UCT research. These included the Wellcome Trust, the National Institutes of Health (NIH) in the US, the European Union, and the Department for International Development (DFID) in the UK, Canada's International Development Research Centre, and the Gates Foundation. The policies were examined in terms of requirements for data deposits and data sharing, as well as time-limits set for these. Findings were that these funders required data deposits, and encouraged data sharing. All except the IDRC required data management plans from researchers, and included time frames for data to be made available to other researchers, some more specific than others.

4 UCT RESEARCH DATA MANAGEMENT NEEDS SURVEY

July 2013

To inform the policy the Project Leader conducted a survey in July 2013 to investigate RDM awareness, needs, and practices among researchers at UCT. An email survey of research projects at UCT, eliciting basic information on research data management planning and processes, and to assist the drawing up of a list of researchers willing to participate in the next stage of the RDMP, in-depth interviews with researchers on how their strategies for dealing with data. The report of the survey is provided as APPENDIX 2.

5 SCOPING STUDY OF RESEARCH DATA MANAGEMENT POLICIES OF JOURNAL PUBLISHERS

July – September 2013

In June 2013 the Project Leader (PL) met with library staff to form a RDM Working Group. The Project Leader gave a presentation on the RDM policymaking and allocated RDM policy information collection activities to working group members as part of several RDM scoping studies to assist policy creation. A Vula site was set up for Scoping Study Teams. Vula is the University of Cape Town's online collaboration and learning environment. It provides a useful space for the Scoping Study Team to communicate and share resources. Resources uploaded include links to university RDM policy documents, relevant standards, and expert organisations. Interesting articles and links are uploaded by team members as they came across them, with comments on their relevance for the UCT project. Team members Maureen Chiware and Linda Kelly undertook a scoping study of RDM requirements of publishers of refereed academic journals. They reviewed 50 of the peer-reviewed journals in which UCT academics published during 2012. Publishing information was obtained from the UCT Research Office. The report of the Journal RDM Policies is provided as APPENDIX 3.

6 SCOPING STUDY OF UNIVERSITY RESEARCH DATA MANAGEMENT POLICIES

April 2013 – February 2014

The Project Leader undertook a scoping study to investigate Research Data Management policymaking at other universities. The study examined the RDM support websites and policy documents of selected universities to identify common issues that need to be addressed to create a workable policy document for UCT.

6.1 STUDY METHOD

Initially universities ranked in the top 50 rankings in the Times Higher Education World University Rankings 2013 were selected for the study. The logic behind this selection was that the more successful universities were thought to adopt strategies to protect and optimise their research data assets, and would thus have established policies for RDM. However, this was not evident from initial findings: 34 of the top 50 universities had some form of RDM in place. This included RDM clauses in existing policies related to research integrity. Some have also created RDM support websites. However, only 13 of the 50 had formal Research Data Management policies. The study was therefore extended to include a number of universities listed as having effective RDM policies on sites of data curation organisations such as the UK's Digital Curation Centre. These included several Australian universities and their inclusion allows better comparison across English speaking countries. Common issues covered by the policies were identified in the study. These items were coded and the spreadsheet of coded items converted to a dataset in the statistical analysis software Stata to allow for detailed analysis.

6.2 THE DATASET

The final dataset covered policies of 29 universities in Australia (7), South Africa (1), the United Kingdom (12), and the United States (9). 13 of these were universities were ranked in the Times Higher Education top 50 universities in the world. 4 were ranked in the top 110 and 12 were not ranked. The earliest policy document was created in 2007, the most recent 2013. The majority of policies (31%) were created in 2011. As the majority of these were UK policies (5 out of the 9 policies created in that year) this could reflect increasing support for research data management from national research funders: In 2011 Research Councils UK issued a set of Common Principles on Research Data which overtly supported funding research data curation activities at UK universities with public funds. This may be an indication of the valuable role funding bodies can play in this regard. Most of the policies examined originated after this date, as universities caught on to the need for proper management of their research assets. The UCT policy initiative is thus timely, and UCT is only the second South African university to create such a policy (the University of Pretoria's Policy for the Preservation and Retention of Research Data was drawn up in 2007).

6.3 STUDY RESULTS

Common issues addressed by policies relate to the why, who, when and how of research data management.

6.3.1 Why manage research Data?

6.3.1.1 Policy purpose

Most policies listed a number of reasons for policy creation, which were to:

- Ensure data assets are identified (20.6% of policies listed this as a reason)
- Ensure data management best practice for research efficiency (65.5%)
- Prevent loss of data (20.6%)
- Meet policy or legislative requirements (13.7%)
- Meet funders' requirements (37.9%)
- Meet journal publishers' requirements (10.3%)
- Justify research findings (72.4%)
- Share data for new research (48.2%)
- Protect Intellectual Property (20.6%)
- Ensure data security (no inappropriate access) (10.3%)
- Allow data citation (3.4%)

- Support the transparent use of public funds (6.9%)

These findings show that policymakers at the universities under review believe RDM will improve research administration and ensure research results can be verified. Some also have funding requirements in mind when developing RDM policy. However, this doesn't seem yet to be an imperative for publication. Protecting university data assets against loss seems less of a concern than protecting IP rights. The relationship between the latter and data security is also not widely recognised. Policies also reflect the low value placed on data citation. This is not surprising as data citation is a fairly new concept, unlike citation of more traditional research materials. Progressive university administrators have taken data citation into account as a policy purpose, however. Possibly they perceive that getting data cited will increase their citation counts and thus their world rankings on scales such as the THE ratings. Increased citations can also be an incentive for researchers to support policy implementation.

6.3.1.2 Data Sharing

Ultimately research data management is aimed at data sharing, with funders or detractors, or with other researchers. The latter will become increasingly important as Open Data initiatives have an impact and funding and academic accolades become linked to data sharing. Data sharing appears to be supported by the conditions in most of the policies reviewed.

Data Sharing Requirements	Freq.	Percent	Cum.
2. Mandates Data Sharing with Provisos	13	44.83	44.83
3. Recommends Data Sharing	4	13.79	58.62
4. Mandates sharing for validation/fund	4	13.79	72.41
5. No Data Sharing Requirements	8	27.59	100.00
Total	29	100.00	

44.8% mandate some form of data sharing, with provisos for data that may have confidentiality or IP requirements preventing sharing. The University of Manchester requires in this case that a non-disclosive version of data be made available. 13.7% of the policies require data to be shared to confirm research outcomes, and to comply with funders' requests for access. 27.5% make no data sharing stipulations. It is recommended that UCT's policy encourage researchers to make their data available for reuse and not only for research verification. The University of Manchester's approach could be adopted for sensitive data or data with licensing provisions. As a final resort useful data with disclosive elements could be deposited with UCT but shared through the University's Secure Data Service. This service, unique in South Africa, shares sensitive data for research under stringent conditions of access.

6.3.1.3 Data Confidentiality

Data confidentiality needs to be covered by UCT's policy even if the policy does not mandate data sharing. Most of the policies examined (51.7%) address this. 3.4% of the policies refer to other policy documents for data confidentiality coverage. 10.3% of the universities deal with this topic on their RDM support websites. However, 34.4 make no mention of data confidentiality in the context of data management. This is surprising, as this may be an issue preventing researchers depositing their data for long-term curation. It is vital that those responsible for curating research data at UCT are aware of the confidentiality issues involved. However, researchers and data administrators may need guidance and training in basic disclosure control to ensure safe data. Educating UCT researchers in this area could be a positive spin-off of the policymaking process.

6.3.1.4 Data Citation

Proper citation will ensure data is discoverable and attributed to the correct source. As mentioned, data citation can be good for university ratings, and can act as an incentive for researchers to curate

their data. Data citation clauses were not included in 62% of the RDM policies reviewed. This confirms the findings from the policy purpose variable.

Data Citation Requirements	Freq.	Percent	Cum.
1. Policy requires data citation	5	17.24	17.24
2. Data citation covered in other docum	2	6.90	24.14
3. Data citation covered on support sit	4	13.79	37.93
4. Data citation not specified	18	62.07	100.00
Total	29	100.00	

Judging from the steady increase in data citations as a result of linked data, it seems wise to encourage the practice in UCT's policy.

6.3.2 Who should be involved?

The "who" of policymaking identifies RDM stakeholders at the university and clarifies their roles. This leads to questions around data ownership (including intellectual property rights) and plans for who should take curatorship of data should Principal Investigators leave the University.

6.3.2.1 Stakeholders

Stakeholders identified in policies include the University's governing body (listed in 24.1% of policies) or Research Administration (41.3%), Academic Directors or Departmental Heads (37.9%), and researchers/PIs (62%). Strangely, 37.9% of policies reviewed did not include researchers as stakeholders, making it difficult to imagine how these policies would be actioned. Only 24.1% included the university's ICT services and the same number listed the University library as a stakeholder. Both these would need to be involved in policy implementation. Research Ethics Committees and Contracts/Grants Offices were identified by only 3.4% of the sample. These Offices as well as the Postgraduate Studies Office should be key stakeholders if RDM policy is to become common practice at UCT. RDM policies need to be seen in relation to policies on research ethics. Any contracts initiated at the university need to be signed off with data management requirements in mind. Research funders can play a vital role by requiring research data management planning from researchers, and by providing grant money for this purpose. So it is disappointing that only 3.4% of the policies mention funders as key role-players.

6.3.2.2 Data Ownership

Data ownership issues will most certainly arise in RDM policymaking. It is vital for researchers to establish ownership of any final data products at initiation of any research project. However, the UCT Research Data Management Needs Survey conducted in July 2013 revealed much confusion around who owns the data underlying UCT research output (UCT). For this reason policies should clearly state the university's position on data ownership and Intellectual Property rights in relation to research data. The scoping study revealed that 58.6% of policies addressed this issue: The University was listed as the data owner (17.2%) or the owner subject to contractual requirements (20.6%) in some cases. The researcher was listed as the data owner subject to contractual or funding agreements in 12.7% of policies. Two of the policies (6.9%) referred to IP policy documents for details on data ownership. This is an option for UCT but it may be better to state UCT's position in the RDM policy document. 41.3% of policies do not address the issue of data ownership. This may lead to questions around the appropriateness of dictating data policy without clear delineation of ownership or at least curatorship of the data to be stored.

Data Ownership	Freq.	Percent	Cum.
1. University Owns Research Data	5	17.24	17.24
2. University owns data subject to cont	6	20.69	37.93
3. Researcher owns data subject to cont	4	13.79	51.72
4. IP policy deals with ownership	2	6.90	58.62
5. Ownership not addressed	12	41.38	100.00
Total	29	100.00	

The draft policy document names the PI as data owner but gives the University curatorship responsibility. This may still need to be debated by administrators at UCT, and clarity on this issue could be one of the beneficial outcomes of the policymaking process.

6.3.2.3 On-going curatorship

Policies need to address data storage and access in the event of the researcher or project PI leaving the university. Some of the university policies reviewed give ownership or at least custodianship of the data to the university under these circumstances: 20.6% of policies mandate that the data remains with the university, 17.2% keep the data but allow researchers to take copies (whatever that means in an era of digital data). A minority (10.3%) permit the original PI to remove the data from the University, but require access to the data wherever it is stored. 51.7% of policies do not address long-term sustainability. This may not be an issue in an era of easy digital data exchange, but the UCT policy should ensure continued access, including for other UCT researchers if data is to be shared.

6.3.3 When should this be done?

6.3.3.1 6.3.3.1 Data deposit time-frames

Some universities have found it useful to prescribe a time-frame for researchers to deposit their data in their institutional repository. 6-7 months after publication was suggested in 6.9% of the policies. 3.4% left deposit time-frames to be determined by the academic discipline concerned, while the majority (89.6%) did not stipulate any time-frame for data deposits.

Data Deposit Time Frame	Freq.	Percent	Cum.
6. 6 Months	1	3.45	3.45
7. 7 Months	1	3.45	6.90
88. Determined by Discipline	1	3.45	10.34
99. Not Stipulated	26	89.66	100.00
Total	29	100.00	

Policymaking here needs to strike a balance between creating a deadline for deposit that is too short and therefore will be ignored by academics, and one that is so long that research is lost in the interim. 18 months to 3 years from publication is suggested as a reasonable time for deposits. More consultation with researchers may be necessary to find a workable time-frame.

6.3.3.2 Data retention time period

Data retention specifications depended largely on the type of data, with longer time prescriptions for clinical data and sensitive or controversial data. Time periods were often seen as dictated by the Discipline involved, or by funding or contractual requirements. Only 7.1% of policies suggested retaining all research data permanently.

dataretention	Freq.	Percent	Cum.
1. 1-5 years	3	10.71	10.71
2. 10-20 years	5	17.86	28.57
3. Permanently	2	7.14	35.71
5. Legal/Funding/Discpline Requirements	12	42.86	78.57
6. Not Stipulated	6	21.43	100.00
Total	28	100.00	

The concept of data life-cycles is changing though, as vintage data is put to new uses and can lead to research breakthroughs. While storage space is not cheap (especially backup storage) it has become more affordable. The future value of research data cannot be assessed in the present. It is therefore recommended that research data collections not available elsewhere be maintained permanently by the University.

6.3.4 How should we do it?

Universities need to make sure policies are in tune with national legislation and other university policies and codes. Policies should provide for research data storage that is interoperable with existing technical infrastructure. They need to ensure data description (metadata) to inform data re-use. Policies also can initiate dedicated RDM support sites to assist researchers in policy compliance. Finally, policies cannot overlook the funding implications of policy operationalisation.

6.3.4.1 Compliance with Other Policies, Codes, and Legislation

Most of the policies reviewed seem to have been created in isolation from existing research integrity policies, or at least these were not overtly considered in the policy documents. Only 10 (34.4%) university policies linked to related policies. RDM policies at these universities were correctly seen as components of research integrity policies or codes, or Open Scholarship initiatives. UCT's policy should be read in the context of the University's Policy for Responsible Conduct of Research and other research-related policies eg Intellectual Property Policies. The UCT administration has also begun discussions around Open Data and it is important that RDM policymaking coordinates with this initiative. Mandating Open Access in the policy document may be premature, but the policy needs to cater for future Open Data requirements at the institution.

While there is a paucity of legislation in South Africa dealing with data for research, UCT's RDM policy needs to be in concert with relevant legislation. The Statistics Act, 1999, and the Protection of Personal Information Act, 2013 should be considered in this regard.

6.3.4.2 Research Data Storage Provision

Successful policy implementation will depend largely on the provision of a well-maintained repository for research data. Data storage at UCT needs to be able to hold large volumes of data in different digital formats. Data security is an issue to be raised here, that is, an optimal means of storing data to prevent inadvertent changes to datasets, prevent unpermitted access, and protect the data against environmental hazards. Institutional storage should be interoperable with existing ICT infrastructure and sustainable in the long-term. It should be user-friendly and allow for self-deposit by researchers. The UCT RDM needs survey showed that UCT researchers have generally stored their research data on personal computers or their own external hard disk drives. Thus a certain amount of cultural change will need to be promoted to encourage deposits in a UCT repository. Ease of use will therefore be important.

55.1% of the university policies examined make provision for data storage on campus, although only 13.7% allow self-deposits by researchers. 6.9% of the universities had not yet allocated storage space for data at the time their policy was created. Surprisingly, 37.9% of policies made no mention of space requirements. Presumably some of the universities concerned were providing storage for their

research data. However, formal commitment to providing a suitable repository to store research data shows a willingness on the part of university administrations to take their role in RDM seriously.

Repository Provision	Freq.	Percent	Cum.
1. Research data repository(self-deposi	4	13.79	13.79
2. Research data repository	12	41.38	55.17
3. Repository being established	2	6.90	62.07
4. Repository not mentioned	11	37.93	100.00
Total	29	100.00	

Research Data Storage options at UCT have been investigated by the Research Portal Project, which is a joint initiative between the Research Office and ICT Services. The project is managed by ICTS with Task Team member Dr Ed Rybicki acting as Academic Liaison to the Project. The outcome of the project was the establishment of a UCT eResearch Portal which went live in February 2012 https://eresearch.uct.ac.za/SitePages/home.aspx The portal provides UCT researchers with tools to support the Research Process, and could be an ideal space for researchers to self-deposit their research output and their research data.

Building a research data repository at the university will cater for researchers whose data does not find a home in a discipline-specific repository, or those in fields where these do not exist. However, UCT's policy needs to make specific allowances for the use of external repositories for research data storage. Many of these data repositories are well established and are preferred by academics as a place to store their data. Just over half of the university policies (51.7%) give express permission for academics to deposit their data in international, national, or discipline-specific repositories. They require an official record to be maintained of any research data stored elsewhere. It is recommended that the UCT research data repository maintain detailed metadata for UCT-generated datasets deposited in external repositories, to ensure the data is traceable and usable.

6.3.4.3 Metadata Requirements

Background and usage information on the research data is a must if the data is to make sense to researchers outside the original project team. Standardised data descriptions (metadata) should be provided with any data deposited in the university repository. Only 8 (27.5%) of the policies in the study stipulated this. However, the absence of policy coverage was dealt with by including metadata information on RDM support websites (27.5% of universities). Only 27.5% of universities covered in the scoping study did not make any stipulations regarding metadata. In 2012 UCT approved a Metadata and Information Architecture Policy to standardise the information that describes, explains, locates, and makes it easier to retrieve, mange or use information resources. A joint UCT Libraries and ICTS workgroup is identifying a range of applicable standards to guide the UCT community.

Metadata	Freq.	Percent	Cum.
1. Policy requires metadata	8	27.59	27.59
2. Metadata discussed on website	8	27.59	55.17
3. No metadata requirements	13	44.83	100.00
Total	29	100.00	

6.3.4.4 6.3.4.4 Research Data Management support and training

Most universities in the study (67.7%) had established support websites to provide researchers and administrators with guidance on data curation. 7.1% were in the process of creating support sites. However, 32.1% did not provide assistance of this kind to their researchers.

Data Support Website	Freq.	Percent	Cum.
1. University has a RDM support site	17	60.71	60.71
2. Support site in development	2	7.14	67.86
3. No RDM support site	9	32.14	100.00
Total	28	100.00	

They sites are key resources to enable RDM. For this reason a team from UCT's library (UL) was established to create a RDM support site (Andrea Walker, Ingrid Thomson, and Michelle Willmers). The website will be managed by the University Library and coordinated by the UCT Director of Libraries. Site access will be from the UL site and the eResearch Portal. The website can provide links to related university policies. International standards exist for the storage and presentation of data files and the creation of metadata, and UCT's site will provide information on these. The submission of Research Data Management plans is fast becoming a requirement to access research funding. Consideration should be given to data management plans for both research groups as well as individuals including postgraduate students, for example, the Supervisor-student MOU should contain a RDM Plan as an addendum. Guidance with this on both the UL site and the eResearch portal will benefit researchers. UCTs RDM site can also link to specific Research Data Management Planning tools such as the University of California's DMPTool and the DCC' s DMPonline.

The RDM support website and eResearch portal will provide researchers with training in managing and sharing their data. Other ways of skilling up researchers and support staff can include formal training in the MLIS Digital Curation Course introduced this year, as well as on-going training workshops. Workshops can be undertaken in collaboration with other research support initiatives. For example, in September 2013 the Project Leader ran a RDM workshop for Postgraduates as part of UCT Library's Savvy Researcher Series which provided attendees with guidance and tools for the creation of RDM plans.

APPENDIX 1:

REPORT ON A UCT RESEARCH DATA MANAGEMENT STAKEHOLDER WORKSHOP

Lynn Woolfrey, UCT RDM Policy Task Team

Date: Monday 24 June 2013
Venue: Boardroom, Room 4.49, School of Economics Building, Middle Campus, UCT
Time: 14h00-16h00

Workshop Participants:
Michael Kyobe, Deputy Dean, Research, Commerce
Jonathan Blackburn (Health Sciences)
Jerome Corns (HS)
Yolande Harley (HS)
Maia Lesosky (HS)
Humanities:
Donald Foster, Deputy Dean, Research, Humanities
Mugsy Spiegel (Humanities)
Colin Tredoux (Humanities
RCIPS:
Andrew Baily (RCIPS)
Piet Barnard (RCIPS)
Warda Sablay (RCIPS)
Andre Le Roux, ICTS (RDM Task Team)
Sakkie Janse van Rensburg, ICTS (RDM Task Team)
Ed Rybiki, Research Office (RDM Task Team)
Gwenda Thomas, Director, UCT Libraries (RDM Task Team)
Martin Wittenberg, DataFirst (RDM Task Team)
Lynn Woolfrey, DataFirst (RDM Task Team)
Research Office: Dianne Bond (RO)
Christina Pather (RO)
Michelle Willmers, SCAP (Co-opted)
Andrea Walker (UL)
Janine Dunlop (UL)
Ingrid Thompson (UL)
Linda Kelly (UL)

1. Background

UCT's Research Data Management Project Plan emphasised that the policymaking process should include broad-based consultation with UCT communities as stakeholders in RDM at the university. In this regard the project milestones included the tabling of a draft RDM policy document at various stages at workshops as well as meetings of appropriate UCT committees. The first Stakeholder Workshop on RDM was held in June 2013.

2. Presentations

2.1 Presentation 1: Research Data Management Policy Requirements

Presented by Lynn Woolfrey, DataFirst/RDM Task Team

This presentation was informed by a RDM Scoping Study initiated in April 2013. The Scoping study included a desk-based study of principles and guidelines related to RDM created by international and national organisations to support best practice. The scoping study included an examination of universities' research data management policies. This covered universities in the English speaking world emulation (Australia, Canada, South Africa, United Kingdom, and United States). Its purpose was to identify common themes in all RDM policies as well as select best practice examples for UCT's policy.

Lynn Woolfrey provided some background to the policymaking process and provided definitions for the term "research data". She discussed motivations for creating policies for the efficient administration and preservation of research data, including mitigating academic fraud, avoiding costly duplication of research and enabling new research based on existing data. She addressed the 6 requirements for a RDM policy common to other universities' policies. These were (i) Defining RDM stakeholder needs and responsibilities, (ii) Providing Secure RD storage, (iii) Catering for data sharing, (iv) Defining time-frames (for data deposits, data sharing and data preservation), (v) Ensuring sustainability of RDM and (vi) Providing training and support. She also discussed the need for UCT to investigate costing models for RDM and allocate long-term funding for this purpose, especially for skilling-up staff to support project implementation.

2.2 Discussion arising from presentation 1

Sakkie Janse van Rensburg discussed a survey he had undertaken of RDM IT infrastructures at Australian universities and said his survey had identified the same issues raised in the presentation. He emphasized the need to address data storage issues. He said funding was available for storage but sustaining future storage needs attention. Finances for this are from a central budget and need to increase. There were suggestions that the ICTS members on the RDM Task Team talk to the DST able long-term funding. The issue of physical and digital resources was raised and Lynn Woolfrey emphasized that the policy was only concerned with digital research data. It was hoped that research data in non-digital formats would be digitized where this was possible, and this would be addressed by the policy. Where this was not possible, e.g. in the case of specimens in health research, repositories would include metadata and links to these, to ensure these are management and discoverable. This would be a second stage of the project though. There was debate around the appropriateness of deposit in a central university repository or a national/ international repository. Lynn Woolfrey stated that where discipline-specific repositories existed and provided optimal storage and access, deposits in these should be encouraged. Some data may need to be stored at the institution as well. But UCT storage would be for research data which did not have a ready home. The institution would require metadata and links to research data stored elsewhere, though.

Lynn Woolfrey emphasized that managing data did not necessarily mean data sharing. However the RDM Task Team was keen to ensure that UCT's policy required the sharing of research data, wherever possible. Participants from the Health Sciences raised the issue of ethics, that is, ways to ensure data security for data confidentiality. It was noted that the RDM policymaking process needed to work with UCT's Open Access initiative, but that this initiative had not yet reached the policymaking stage.

Issue of who owns the data produced by researchers at UCT was raised. If the data was housed offsite then probably UCT would not have ownership of the data. Participants felt that with regard to questions of ownership there may be a need to keep all research data at UCT, even if this is already deposited with another repository. That did not mean sharing of data would be via UCT.

2.3 Presentation 2: An overview of funders' Research Data Management requirements

Presented by Warda Sablay, UCT Research Contracts and Intellectual Property Services (RCIPS)

Warda Sablay presented an overview of RDM requirements of key funders of UCT research. These included the Wellcome Trust, the National Institutes of Health (NIH) in the US, the European Union, and the Department for International Development (DFID) in the UK, Canada's International Development Research Centre and the Gates Foundation. She examined their policies in terms of requirements for data deposits and data sharing, as well as time-limits set for these. She also assessed whether researchers were required to submit data management plans with project proposals. Her findings were that all the above funders required data deposits, as well as data sharing. All except the IDRC required data management plans from researchers, and included time frames for data, some more specific than others.

2.4 Discussion arising from presentation 2

Current data management models at faculty and departmental levels at UCT were discussed. Resistance from academics to formal research data archiving structures was raised by Mugsy Spiegel and he suggested investigating this further to ensure the project is appropriate to ensure its success. Lynn Woolfrey explained that she was working on a survey instrument for a Research Data Management Needs Survey to be completed by UCT researchers to be undertaken in July 2013. She will upload the draft survey questionnaire as a Google doc for comments from workshop participants, with a time-frame for this feedback. Gwenda Thomas will arrange for the RO to send out the survey. She will also contact the Deans and the Faculty Research Committees, and Departmental Heads, to ensure support for the survey. Participants agreed that ideally the survey should be conducted in mid-July to obtain results at the end of July, perhaps in time for the budget meeting on 18 August 2013.

APPENDIX 2:

REPORT ON A UCT RESEARCH DATA MANAGEMENT NEEDS SURVEY

Lynn Woolfrey, UCT RDM Policy Task Team
July 2013

1. Background

In 2013 UCT Research Committee embarked on a project to create a Research Data Management Policy for the university. The aim of the policy is to join the ranks of top universities worldwide in creating model research data management practices for their researchers. The policy would endorse principles of sound research data management, identify stakeholder roles and responsibilities, and establish mechanisms for the secure long-term preservation of the data underlying research at UCT. The policy will ensure the data is stored in a manner that fulfils any legal, statutory, ethical, and contractual requirements, and supports UCT's research integrity policies.

The proper handling of evidence for research output will prevent the loss of valuable and potentially reusable data. It will also protect the university and researchers by allowing verification of research results in the case of suspected scientific misconduct. Policy guidelines aim to protect the university's intellectual property rights while at the same time allowing appropriate access to suitably anonymised data for further research collaborations. Research sponsors are increasingly requiring data management or data sharing plans in funding requests. Journal editorial boards have begun to demand the underlying data with journal submissions. The policy will assist researchers to comply with these funding and publishing requirements while relieving researchers of some of the burden associated with these stipulations.

2. Survey Method

To inform the policy the Project Leader conducted a survey in July 2013 to investigate RDM awareness, needs, and practices among researchers at UCT. The survey questionnaire was sent via email from the UCT Research Office to all researchers at the university. Data was collected on the respondent's position at UCT and on their research areas of interest. Questions on the researcher's data were concerned with data sources, ownership of the data, data formats, size of data holdings and perceived importance of the data for on-going research. Research data management practices were investigated, including allocation of responsibility for the data, data storage, backup and retention, and whether researchers or their research projects had any formal data management plan in place.

3. Survey Results

3.1 Research Communities Represented

The survey returned 145 responses which was a fairly rewarding return given researchers' lack of enthusiasm for taking surveys. The majority of respondents were academic or research staff (80.3%), with the remainder being visiting researchers, research administrators or postgraduate students. Research fields covered the spectrum of academic departments with Health Sciences providing the most returns. This may reflect greater concern among health researchers for solutions to handling their research data, or more stringent data handling requirements of funders in the field. Another explanation for this could be that health researchers have been exposed to concerns around data management issues such as data confidentiality and security when dealing with patient records.

1. Please indicate your position at the University of Cape Town.

	Response Percent	Response Count
Postgraduate research student	4.1%	6
Academic/research staff	**80.3%**	**118**
Visiting researcher	0.7%	1
Research administrator	6.8%	10
Other	8.2%	12
If you selected Other above, please specify.		11
answered question		**147**
skipped question		**1**

Table 1 Respondents

3.2 The Research Data

Respondents were asked from where they sourced the data they use for their research. Most (68.3%) obtained their data as raw data from instruments (including from questionnaires used in surveys). 35% generated data from a programme, 27.5% from laboratory notes and 24.2% used patient data. A surprisingly large amount of respondents (53%) are working with qualitative data, although all but a few (5.9%) of these also used other data types. Data formats also spanned the range, with most researchers using text documents (82.5%) and spread-sheets (77.5%). A table is provided below with a breakdown of data formats (percentages do not add up to 100% as respondents could choose all sources that applied).

5. Please select the data formats that apply to your research or project.

	Response Percent	Response Count
Data in a database (e.g. MySQL, Oracle)	34.2%	41
Images, scans or x-rays	40.8%	49
Digital audio	24.2%	29
Digital video	25.0%	30
Text documents (e.g. Word, PDF)	**82.5%**	**99**
Spread-sheets (e.g. Excel)	77.5%	93
Raw data in American Standard Code for Information Interchange (ASCII)	10.8%	13
Raw data is a proprietary software format such as SAS/SPSS/Stata	28.3%	34
Geospatial data - vector (e.g. co-ordinate lists, CAD files, shape files, geo-databases)	8.3%	10
Geospatial data - raster (e.g. scanned maps, satellite imagery, aerial photography)	8.3%	10
Other	8.3%	10
	If you selected Other above, please specify.	11
	answered question	**120**
	skipped question	**28**

Table 2: Data Formats used by UCT Researchers

Respondents were asked to give the size of their current data holdings. Data collections ranged from 100 megabytes to 10 Petabytes. Approximately half of the respondents (50.51%) had no idea of the size of their data holdings, which is an indication that they have perhaps not given much thought to data storage needs going forward.

3.3 Research Data Curation Practices

63% of respondents rated the research data they held as vital for their research, meaning they would not be able to continue with their research if the data were lost. 32.8% regarded their data as important and only 4.2% saw their data as ephemeral. In contrast to the high value they accorded their data, respondents' data management practices were not uniformly sound. Data stewardship roles were generally clearly allocated: Responsibility for managing the data rested largely with the respondent-

researcher (86.4%). Where this was not the case a research assistant (30.5%), a designated person on the project (28.8%) or a Project Manager (18.6%) were allocated this task. Surprisingly, even for data to be curated for more than 10 years, not much use was made of official data archives set up for this purpose: Only 2.5% of respondents deposited their data with a national data archive, 5.9% with a discipline-specific national data archive, and (3.4%) used discipline-specific international data archives.

8. Who is currently responsible for managing the data? (Please select all that apply.)

	Response Percent	Response Count
Project manager	18.6%	22
Designated person on project	28.8%	34
External project partners	15.3%	18
IT staff within the department	5.1%	6
UCT IT staff	1.7%	2
Research assistant	30.5%	36
Yourself	**86.4%**	**102**
National data archive	2.5%	3
Discipline-specific national archive	5.9%	7
Discipline-specific international archive	3.4%	4
Nobody	1.7%	2
Don't know	0.8%	1
Other	11.0%	13
	If you selected Other above, please specify.	14
	answered question	**118**
	skipped question	**30**

Table: 3 Research Data Manager

Data retention periods of responding researchers were well within the international recommendations: 5.8% aimed to keep their research data only until the Project was completed, but 21.8% would retain their data for the internationally accepted minimum retention period of up to 5 years. Nearly a quarter (23.5%) of respondents would retain their data for up to 10 years, and 37.8% envisaged keeping their data for more than 10 years. However, a significant minority (10.9%) did not know how long they would keep their data.

The provision of data storage is a key component of any university RDM initiative. Questions in the survey around where UCT researchers store and backup their data provide feedback to the policy team on researchers' data storage needs. Although more than a third of respondents (37.8%) aimed to retain their data for more than 10 years, almost half of respondents (42.7%) were storing their data on external hard drives. This is suitable for storage during the life of the project, but long-term data would ideally be entrusted to a secure repository accessible to all project members. However, institutional repositories were not routinely used for long-term storage: UCT servers managed by the university's Information Communication and Technology Services (ICTS) were used for data storage by 6.8% of respondents, and departmental servers by 11.1%.

10. Where are the data stored?

	Response Percent	Response Count
UCT ICTS server	6.8%	8
Departmental server	11.1%	13
CDs/DVDs	2.6%	3
USB/Flash drives	7.7%	9
External hard drives	**42.7%**	**50**
Tapes	0.9%	1
Third party/ Cloud/ Commercial data service	7.7%	9
Don't know	0.9%	1
Other	19.7%	23
If you selected Other above, please specify.		35
answered question		**117**
skipped question		**31**

Table 4: Data Storage

Research data is updated regularly by respondents, mostly daily (21.1%) weekly (23.7%) or monthly (21.9%). 7.0% of respondents were updating their data annually. A surprising 16.7% of respondents did not know if the data was being updated. This may reflect confusion about whether updates refer to additional findings added to a database or data checking. Most researchers (75.2%) back up their data, mostly on external hard drives (72.7%) or with third-parties commercial data/cloud services (21.6%). Once again UCT servers were not the repositories of choice: Only 8% used UCT servers managed by ICTS for backups and 11.4% made use of servers in their departments for this purpose. An alarming 18.8% did not back up their data and 6% didn't know if the data was being backed-up.

3.4 Research Data Management Planning

Only 14.3% of respondents currently have a formal research data management plan for their research project. 79% do not have any formal plan and 6.7% did not know if a plan existed for their projects' research data. This indicates that it is time for UCT to establish policy and procedures to enable researchers to manage and store their data in a manner that ensures they can find and use it again, for further research or collaborative research, or to justify original findings.

14. Do you currently have a formal Research Data Management Plan in place for your research/project?

	Response Percent	Response Count
Yes	14.3%	17
No	79.0%	94
Don't know	6.7%	8
answered question		119
skipped question		29

Table 5: Research Data Management Plans

4. UCT Researchers' Data Management Needs

4.1 Spreading the Burden of Data Stewardship

From comments provided by respondents in the survey it was clear that some disciplines were already actively managing their research data – Astronomy is an example, where respondents reported that the South African Astronomical Observatory (SAAO) is engaging with research data management issues. Health researchers are also attempting to curate their research output more effectively. Positive comments were received about the Project's aims, asserting that research data was "a valuable heritage" that should be preserved. Several respondents felt there was a strong need for better curation of research data at Departmental level, but could see the advantages of undertaking this as a component of university-wide practice.

Specifically, researchers welcomed assistance with creating the research data management plans required to access funds from some agencies. Many of them also stated that they would make use of any free secure storage provided by the university. A few had lost data through hardware crashes or had "run out of space". Some felt that data collaboration possibilities could be increased through making research output available to other researchers on a shared research data repository.

On the other hand many respondents felt the task to be overwhelming. A fairly common response from researchers was that their research data was immune to any sort of ordering, and there was "no formal way of dealing with the mass of [their] research data". They were also concerned that the demands of RDM would add to their already heavy work-loads, and "increase bureaucratic restrictions on researchers". This is a very real concern. Cases exist where sound policies and secure infrastructure for RDM have been established by universities but there has been no uptake by researchers. The time and effort required to find and collate their data for deposit has been a strong disincentive to participation. In these cases support and training for RDM have been under-utilised and repositories have remained empty (Nelson: 2009).

RDM may seem daunting to researchers. However, no-one's data is uniquely unmanageable. Successful RDM repositories and practices have been established at universities worldwide. Difficulties around curating research data have been dealt with at universities by allocating resources and staff to support RDM. This takes some of the odious task of data preparation and data deposit off the shoulders of researchers. A scoping study of RDM at the top 50 universities listed in the Times Higher Education World University Rankings 2012-2013 undertaken by the RDM Policy project confirmed that 31 of these universities have some form of RDM commitment and support, and 21 have the full suite of RDM tools, in the form of policy documents, web-guides and expert staff (RDM scoping study, 20131). Effective, Open Source RDM tools already exist and continue to improve, as the practice becomes more entrenched. RDM introduces researchers to data organisation throughout the lifecycle of the data, from initial data collection through to deposit of data in a trusted repository along with the final published output. Planning for data handling at project onset not only makes the task more manageable, it can be invaluable in assisting the Research Process. UCT will need to skill up research and support staff to ensure they are not overwhelmed by RDM requirements.

4.2 Dealing with Concerns around Research Data Sharing

The goal of all university RDM policies is to increase research efficiency through introducing mechanisms to identify and preserve data assets. These policies ensure compliance with funding requirements to store data for specified periods to enable review where research findings are questioned. Well curated data is also reusable by original research teams to extend their work (RDM Scoping Study). An additional reason to preserve and manage research data is to enable reuse for purposes other than the original research aims. Data sharing is an additional funding requirement of certain grant-makers and a condition of acceptance for publication in major journals (RDM Scoping Study). South Africa's National Research Foundation has recently introduced a data management plan requirement for research funding proposals, but this does not stipulate data sharing. Thus many local researchers are not comfortable with the idea of curating their data for reuse by non-collaborating researchers.

4.2.1 Ensuring Appropriate data reuse

One survey respondent raised the need for Open Access requirements to be taken into account in the policymaking process. However, many respondents felt that their research data would not be useful to anyone else, or feared their data would be misinterpreted by other researchers. This concern was voiced in comments such as "no-one is capable of using my raw data, however well I supply the metadata". While this may be more perception than reality - after all, if the data is too complicated for anyone else to understand, then it is too complicated for peer review, which raises problems with research integrity- this is an issue the policy should cover. Academics can benefit from collaboratively using other researchers' data. Data usage statistics from our Data Service at UCT show that even the most obscure dataset can be useful and that the value of the initial research data may increase if it is combined with new data sources. To allay fears of data misinterpretation research data stewardship at UCT would of necessity include assisting researchers to describe their data to ensure appropriate future usage.

4.2.1 Dealing with Confidential Research Data

One respondent stressed that "data management raises serious ethics issues." This is true when the data is to be curated for reuse. Legal and ethical considerations around confidentiality should be addressed in the final policy document and policy implementation. Most RDM policies acknowledge that there are some data sources that cannot be shared ethically but suggest public-use sub-sets of the data could be made available. Another option with regard to sensitive data is to house and make the

[1] This scoping study is on-going and has been undertaken by the RDM Project Leader.

data available in a secure supervised environment. This is the practice in major universities worldwide, where sensitive government and research data is provided to academics in secure spaces by secure data services. UCT has a Secure Data Service, established by DataFirst in 2012 in the School of Economics building. Deposits with this service could allay fears around data confidentiality

4.2.2 Data Ownership Issues

The idea of "centrally-managed" research also filled some respondents with trepidation about losing control of their work. Researchers fear that their work may be overshadowed by better analysis of their data by others. However their 18 month embargo head-start and their intimate knowledge of the data usually ensure them a publishing advantage. The policy approach to these concerns around data sharing could be twofold: (i) A gradual introduction of the data sharing component and (ii) Open Data advocacy by the Research Office and Library. In the initial stages RDM at UCT should ensure researchers are comfortable with research data management for their own purposes rather than for data sharing. Most university RDM policies have dealt with the issue in this manner. They provide guidance for data sharing where this is a funding requirement, and encourage data sharing, but do not mandate it. UCT's policy could similarly recommend data sharing for the good of science, and ensure the final policy is compatible with any future Open Data policies which may be introduced at UCT in the future.

Open Data advocacy could centre on RDM incentives for researchers. There are many ways researchers can benefit directly from curating data for reuse. For example, preparing data from project outset requires careful documenting of data work which can reduce the chances of the researcher making data analysis errors and thus avoid them the embarrassment of errors in their work being identified post-publication. Data citation can also increase a researcher's profile. Henneken and Accomazziins, 2011 provide some evidence to confirm that sharing data increases citations for researchers, at least in some fields. This can be a big selling point for data sharing, particularly for early-career researchers. These advantages will become evident only once a critical mass of researchers makes use of any RDM infrastructure and support put in place at UCT. Policy implementation should include easy self-deposit solutions, and other forms of support for researcher participation.

4.4 Other Concerns

Institutional records management imperatives were also mentioned. While the management of institutional records used in the administration of the university is outside the ambit of the research data management project there are issues of common interest which can be informed by the current project work.

APPENDIX 3:

REPORT ON RESEARCH DATA MANAGEMENT: A REVIEW OF JOURNALS

Linda Kelly & Maureen Chiware
With acknowledgements to Dianne Bond, UCT Research Office
August 2013

1. Brief summary

1.1 Overview

The websites of 50 journals were reviewed to determine the number of journals that specified or adhered to a research data management policy (RDMP). These were journals with the highest number of UCT authored articles published during 2012.

The review took the following variables into account:

- RDMP policy or statement
- Storage requirements for research data
- Sharing requirements for research data
- Time-frames associated with research data

1.2 Findings

The following graphs indicate the findings of the review for the above variables.

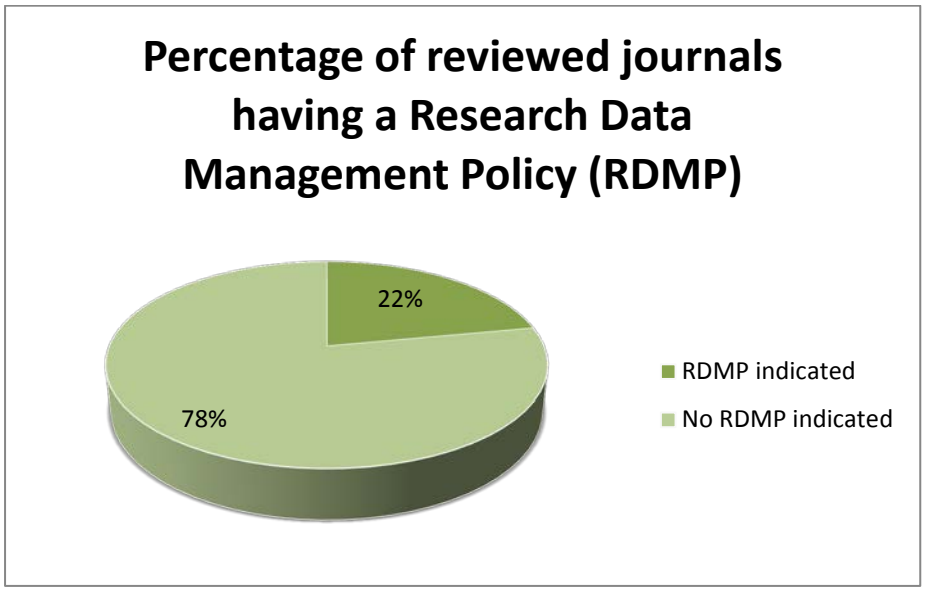

Figure 1: Percentage of reviewed journals having a RDMP

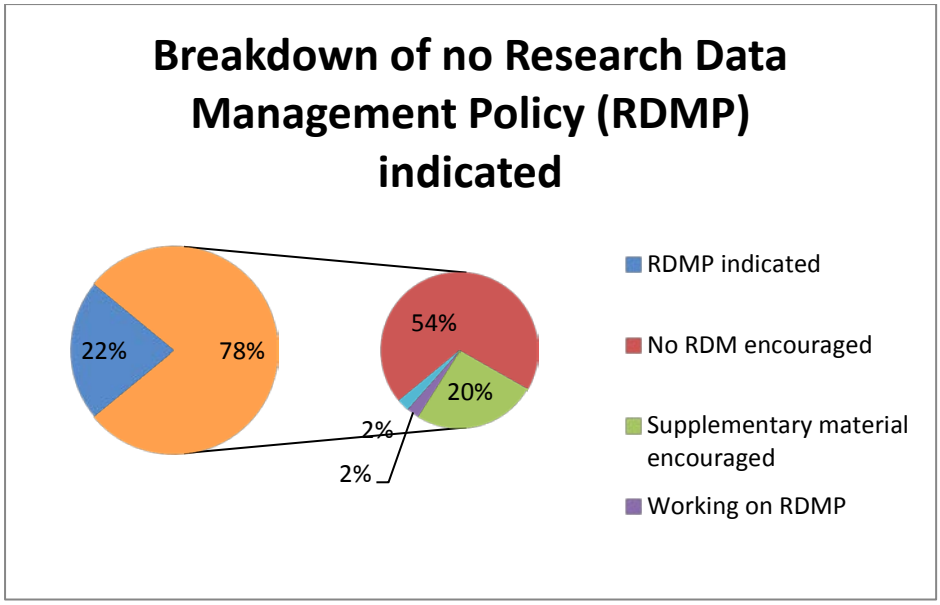

Figure 2: Percentage of reviewed journals without a RDMP

Figure 1 indicates that 22% had some form of RDMP. The breakdown of journals that did not indicate a set RDMP (depicted in Figure 2) shows that 10 journals (20%) encouraged or allowed authors to submit 'supplementary material'. One journal (2%) indicated that they were working on a RDMP while another journal (2%) specified agreements with funder bodies about depositing research data. Therefore, a total of 42% of the journals either indicated some form of RDMP (22%) or encouraged supplementary data (20%).

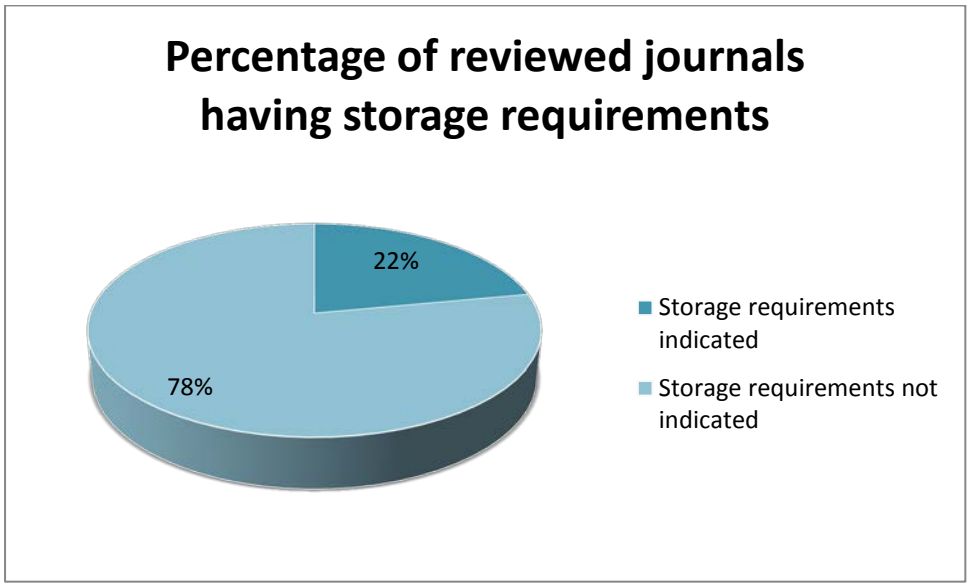

Figure 2: Percentage of reviewed journals having storage requirements

Various storage requirements associated with research data were indicated by 22% of the journals reviewed. This included capping the amount of research data and recommending file formats.

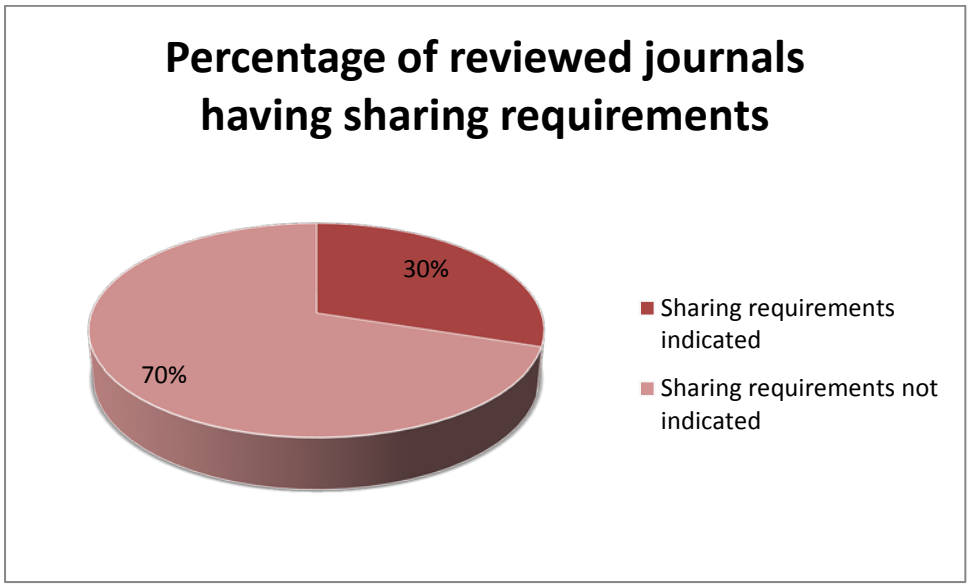

Figure 3: Percentage of reviewed journals having sharing requirements

Nearly a third (30%) of the reviewed journals had requirements relating to the sharing of research data. This included submitting the data with the article and/or in a specified repository or database. This value is higher than the journals that indicated a RDMP (22%) but less than the sum of journals that indicated a RDMP and encouraged supplementary materials (22% + 20% = 42%).

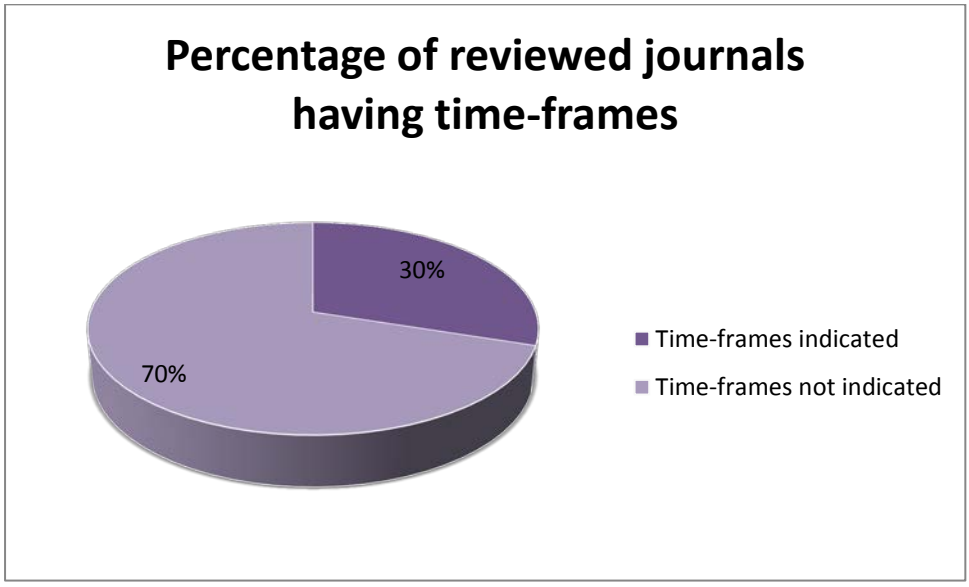

Figure 4: Percentage of reviewed journals having time-frames

Figure 4 shows that 30% of the journals reviewed had some form of time-frame that related to submitting the research data. As stated above, this value is higher than the journals that indicated a RDMP (22%) but less than the sum of journals that indicated a RDMP and encouraged supplementary

materials (22% + 20% = 42%). It is this supplementary material that also has various forms of time frames such as submitting the research data with the article, how long the data should be available in a repository or database and by when research data in a repository should be deposited and linked to the article.

2. Expanded report

2.1 Introduction & Aim

The aim of the review was to determine the number of journals that specified or adhered to a research data management policy (RDMP). These were journals with the highest number of UCT authored articles published during 2012.

2.2 Methodology

2.2.1 Sample size

All articles published during 2012 by UCT authors were provided by Dianne Bond from the UCT Research Office. These articles were then sorted alphabetically according to their journal title and the number of articles per journal was calculated in the Excel spreadsheet entitled 'RDMP_Review of Journals' (see tabs 1-4 for this analysis). This data was used to rank the journals according to the number of articles that had been published in 2012 (see tab 4 "RDMP review"). The 53 journals with the most articles were identified.

The following three journals were excluded from the review, decreasing the number to 50 journals:

Journal	Reason for exclusion
Journal Of High Energy Physics	Ulrich Periodicals Directory states that publication ceased in 2000
Acta Juridica	No website was located for journal
Cochrane Database of Systematic Reviews	No information obtained because website page loads incorrectly

Table 1: Journals excluded from the review

2.2.1 Analysis: Review

Each journal was reviewed by searching its website for information related to research data requirements or policy. Ulrich's Periodicals Directory was used to to obtain the website of each journal. As stated above, the variables in Table 2 where into account when reviewing each journal's website.

Variable	Explanation
RDMP policy or statement	A statement or policy about research data that authors must adhere to when submitting their article
Storage requirements for research data	Storage requirements such as file size or type
Sharing requirements for research data	Sharing requirements about where and how the research data would be accessed
Time-frames associated with research data	An indication of when must the research data be published

Table 2: Variables used for the review

2.2.2 Limitations

It was noted that unknown factors could impact the occurrence of policy or statements about research data management which were not taken into account due to time constraints. For example, the journal

websites could display information that may not reflect current practices or policies. For the purpose of this study, it was assumed that if a RDMP was not stated, it did not have one.

The authors acknowledge that the findings of this study should be validated and unpacked by further research in order to understand current practice and trends related to research data management.

2.2.3 Findings

See section 1.2 Findings above.

2.2.4 Summary

The review has indicated that while about a quarter of the journals reviewed had some form of research data management policy, a further 20% encouraged or allowed authors to submit supplementary research data with their article. These were 50 journals that had the most UCT authored articles published in 2012.

APPENDIX 4: UNIVERSITY OF CAPE TOWN RESEARCH DATA MANAGEMENT POLICY SCOPING STUDY

Codelist for Variables

Country

1. Australia
2. South Africa
3. United Kingdom
4. United States

University

1. Brunel University
2. Cornell University
3. Duke University
4. East London , University of
5. Edinburgh University
6. Edith Cowan University
7. Essex, University of
8. Harvard University
9. Hertfordshire, University of
10. Imperial College London
11. Johns Hopkins University
12. La Trobe University
13. Manchester, University of
14. Melbourne, University of
15. Monash University
16. New York University
17. Newcastle, University of
18. Northampton, University of
19. Northwestern University
20. Oxford University
21. Pretoria, University of
22. Queensland University of
23. Sheffield, University of
24. Southhampton, University of
25. Stanford University
26. Sydney, University of
27. Tennessee, Knoxville, University of
28. Warwick, University of
29. Wisconsin-Madison, University of

Times Higher Education World University Rankings 2013 score

0. Not ranked
1 – 110

Policy implementation date

2007
2008
2009

2010
2011
2012
2013

Policy purpose

1. Ensure data assets are identified
2. Ensure data management best practice for research efficiency
3. Prevent loss of data
4. Meet policy or legislative requirements
5. Meet funders' requirements
6. Meet journal publishers' requirements
7. Justify research findings
8. Share data for new research
9. Protect Intellectual Property
10. Ensure data security (no inappropriate access)
11. Allow data citation
12. Support the transparent use of public funds

Data sharing

1. Mandates research data sharing
2. Mandates data sharing with provisos
3. Recommends data sharing
4. Mandates sharing for validation/funders
5. No data sharing requirements

Data confidentiality requirements

1. Policy mandates data confidentiality
2. Other data confidentiality policies
3. Website deals with confidentiality
4. Confidentiality not mentioned

Data citation (attribution)

1. Policy requires data citation
2. Data citation covered in other document
3. Data citation covered on support site
4. Data citation not specified

Stakeholders

1. University/ Governance Body
2. Central Administration/ Director
3. Research Administration/ Director
4. Research Committee
5. Research Ethics Committee
6. Grants/Contracts Office

7. Academic Directors/Departmental Heads
8. Researchers/ Principal Investigator
9. ICT Services
10. Library/information Services
11. Library Committee
12. Records/ Information Management Services
13. EResearch Centre/ Digital Curation Centre
14. Research Funders
15. Not listed

Data ownership and Intellectual Property rights

1. University owns research data
2. University owns data subject to contracts
3. Researcher owns data subject to contracts
4. IP policy deals with ownership
5. Ownership not addressed

On-going curatorship

1. University keeps the data if PI leaves
2. University keeps the data, PI may take a copy
3. PI takes the data but University has access
4. Does not cater for PI's departure

Data Deposit Timeline (months)

6
7
88. Determined by discipline
99. Not stipulated

Data Retention Period (years)

1
2
3
5
10
15
20
44. Permanently
55. Retain until student graduates/leave the university
66. Determined by legal requirements
77. Determined by funder
88. Determined by discipline or community
99. Not stipulated

Compliance with other university policies and codes

1. Mentioned
2. Not mentioned

Compliance with national legislation

1. Mentioned
2. Not mentioned

Research Data Repository provision

1. Research data repository with self –deposit options
2. Research data repository
3. Repository being established
4. Repository not mentioned

Caters for storage in an external (national or subject-specific) repositories

1. Yes
2. No/not mentioned

Data Description (Metadata) Requirements

1. Policy requires metadata
2. Metadata requirements discussed on website
3. No metadata requirements

Research Data Management Support Site

1. University has a RDM support site
2. Support site in development
3. No RDM support site

Research Data Management Plans

1. Policy mandates research data management plans from researchers
2. Policy recommends research data management plans from researchers
3. Research Data Management plans covered on support site
4. RDM plans not mentioned

Funding for RDM

1. University will help cost RDM but no funding offered
2. Suggest RDM Funding should be sourced from project funders
3. Funding not dealt with in the policy document

Lynn Woolfrey
UCT RDM Policy Task Team
March 2014

UNIVERSITY OF CAPE TOWN DRAFT RESEARCH DATA MANAGEMENT POLICY

Policy name: University of Cape Town Research Data Management Policy

Responsible Executive:Dean of Research, University of Cape Town

Responsible Office: Research Office

Issued: 20140131

Version: Draft Policy Document Version 2

Document URL:

1. Introduction

This policy states the reasons why it is important to manage the data underlying research output, and allocates responsibilities and enables procedures for managing research data at the University of Cape Town (hereafter referred to as "the University").

2. Policy Purpose

The purpose of this policy is to ensure consistent, university-wide practice related to managing research data. Such practice will support research efficiency and research integrity at the University, assist researchers to comply with legal requirements, and funding and publishing stipulations

2.1 Support for better research

Research data is the evidence base for research undertaken at the university and forms part of the university's intellectual assets. The sound management of research data during research projects and in the long term is good research practice that promotes academic rigour and ensures that the university's data assets are identified and recorded and preserved to:

- Prevent the loss of valuable data assets
- Protect the university's intellectual property
- Support research transparency

- Ensure research reproducibility in defence of research findings
- Prevent duplication of research
- Ensure data security
- Allow data reuse for new collaborations and to extend research findings

2.2 Compliance with legislation and international standards

This policy will support the requirements of international codes of best practice and national legislation concerned with data handling, and related University of Cape Town policies.

2.3 Compliance with funders' and publishers' requirements

Funding agencies and publishers of peer-reviewed journals recognise that active Research Data Management is essential for sound research, and require research data management plans to be submitted with research proposals. This policy will assist researchers to meeting these research data management planning requirements.

2. Definition of Research Data

For purposes of this policy, "Research Data" means all records necessary for the reconstruction and validation of research findings, in digital format. Where possible, research evidence which is not "born digital" should be digitised for long-term preservation.

3. Stakeholder roles and responsibilities

The University's Research Administration is responsible for initiating this policy and ensuring compliance with its terms. The Research Administration will allocate funding for research data management at the University, including for data storage, advocacy and training of administrators and researchers.

The University's Information and Communication Technology Services (ICTS) will be responsible for providing and managing research data storage.

The University Library Director is responsible for ensuring library staff provides information, training and support to researchers on research data management. This includes responsibility for the establishment and maintenance of a research data management website and guides for researchers.

Deans, Departmental Heads, research administrators and grants and contracts personnel should be aware of this policy and support researchers in its implementation. Principal Investigators are responsible for maintaining and preserving research data during the life of the research project.

Principal Investigators are responsible for depositing research data with the University's repository or an external repository on project close-out.

4. Data ownership

Research output and research data belong to the researcher or project producing the research, unless funding agreements specify otherwise. However the University has stewardship over the research data and has an obligation to ensure this data is stored and preserved in the long term.

5. Research Data Storage

The University's research administration undertakes to provide secure research data storage in the form of an institutional repository which provides for self-deposit by researchers. They will also

support the deposit of research data in established national or international and discipline-specific repositories. Any research data deposited in an external repository should be registered with the University to ensure the data is traceable. Research data should be preserved on an on-going basis as the value of this data may only be realised in the future. If the researcher leaves the University they may take their research data. However a copy of the data must remain with the University which has ongoing custodial responsibility for the data. This responsibility will be transferred to the relevant academic department if the principal investigator leaves the University.

6. Data formats and metadata

Deposited data should be in usable formats, preferably those prescribed by international data curation standards. All research data deposited in the University repository or external repositories should have descriptive information (metadata) on how to access and use the data. This metadata should be created according to international standards.

7. Data sharing

Sharing research data ensures academic rigour and helps to advance science. Research data must always be available to project collaborators and research funders. Data should be shared more widely where this is a stipulation by research funders. Where sharing is not a funding requirement, the University encourages its researchers to make their data available to other researchers within 18 months of publication. Co-authorship should not be a requirement for sharing the data with other researchers. Rights to reuse or publish research data should not be handed over to commercial entities without the researcher retaining the rights to make the data available to other researchers for re-use. Research data should be shared in a manner which safeguards data confidentiality and the privacy of participants.

8. Data citation

Researchers using third party research data should acknowledge the sources of their data. Data citation recognises the contribution made by researchers who generate and share research data.

Related policies:

UCT research policy http://www.researchoffice.uct.ac.za/research_integrity/policies/uct_research/

UCT policy for responsible conduct of research

http://www.uct.ac.za/downloads/uct.ac.za/about/policies/UCTresearchconductpolicy.pdf