UNIVERSITY OF CAPE TOWN

RESEARCH DATA MANAGEMENT POLICY

Policy name: University of Cape Town Research Data Management Policy

Responsible Executive: DVC (Research & Internationalisation)

Responsible Office: Research Office

Issued: 17 March 2018

Version: Draft Policy Document Version 4- revised

Document URL:
http://www.uct.ac.za/sites/default/files/image_tool/images/328/about/policies/TGO_Policy_Research_Data_Management_2018.pdf

A. POLICY STATEMENT

1. Introduction

The drivers and principles for managing research data at the University of Cape Town (hereafter referred to as "the University"), are defined in response to the FAIR principles of Open Science as a series of research practices related to the increasing use of digital infrastructure.[1] The benefits of Open Science are recognised in the wider dissemination of research outcomes; improved

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

opportunities for collaborative research; improved responsiveness to societal challenges; the leveraging of data-intensive science as an economic driver; and greater research integrity.

In addition, an increasing number of policies published by funders of research to ensure the validation of research results; to provide research opportunities in data reuse, and to enable actionable and socially-beneficial science to address global research challenges. In addition, recent emphasis on the principle of Open Access by default to data resulting from publicly –funded research requires consideration of the necessary limits on openness, particularly relating to personal information and commercial considerations to assist researchers to comply with legal requirements, and emerging terms of funding and scholarly publishing.

This policy is defined in accordance with relevant standards and community best practice in the international context, as outlined by the International Council of Science Unions (ICSU) and CODATA.[2]

2. Purpose Statement

The purpose of this policy is to transform the way research is conducted at UCT by accelerating discovery, increasing the value of research decision-making, and catalysing changes throughout the economy and society that are of value to all citizens.

The University seeks to ensure consistent research practice related to data management principles that support effective data sharing, including open access; and the need for data to be discoverable, accessible, reusable and interoperable to specific quality standards.

2.1 Data Principles

The following common principles of research data management provide a transparent and coherent framework to this policy:

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
- Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.

[2] (2014) Hodson, S. and Molloy, L. *Current Best Practice for Research Data Management Policies*. Report on Current Best Practice for Research Data Management Policies commissioned from CODATA by the Danish e-Infrastructure Cooperation and the Danish Digital Library and submitted in May 2014. (https://zenodo.org/record/27872#.VopzkJD8Lcc)

- In terms of the Intellectual Property Policy research data are owned by the University unless otherwise agreed to by the University in terms of a research contract concluded Fwith a funder.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data.
- Published research results should include information on how to access the supporting data.
- UCT recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, associated policies, guidelines and practices ensure that these are considered at all stages in the research process.
- To ensure appropriate recognition for the effort involved in collecting and analysing data, researchers may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The time frame may be determined by the funder and should not exceed a maximum period of two years following the date of publication.
- In due recognition of the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for scholarly communications should be both efficient and cost-effective in the use of public funds.

3. Definitions

3.1 Definition of Research Data

For purposes of this policy, the definition of scientific research provided by the *OECD Principles and Guidelines*[3] serves both the Human and Natural Sciences:

""research data" are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the

[3] (2007) OECD Principles and Guidelines for Access to Research Data from Public Funding. (http://www.oecd.org/sti/sci-tech/38500813.pdf)

scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated."[4]

The OECD definition specifically excludes "laboratory note books, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice)."[5]

3.2 Definition of Terms

Creative Commons [6]means a non-profit organisation which is committed to facilitating the legal sharing of creative works though a range of licences which allow creators to stipulate which rights they reserve, and which rights they waive for the benefit of other creators. Creative Commons licences follow a "some rights reserved" model in contrast to traditional copyright, which follows an "all rights reserved" model. Creative Commons therefore provides a continuum of rights between "all rights reserved" on the one end of the continuum and "no rights reserved" (public domain) on the other end;

DMP means data management plan, a formal document that outlines data will be handled both during the research project, and after the project is completed.

Data curation means the active management of data through its lifecycle to enable discovery and provide for reuse over time to scholarly and educational activities across the sciences, social sciences, and the humanities;

Data management means the administrative process by which the data is acquired, validated, stored, protected, and processed throughout its lifecycle; and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users;

Metadata means structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (NISO, 2004). Examples of key metadata elements are: Title of the article/document; Author (Creator) of the document; Description of the content; Source of the document, Date created;

Open Access means the immediate, online, free availability of research outputs that it can be accessed by anyone in the world and is free of most copyright and licensing restrictions.

[4] OECD Principles. p.13.
[5] OECD Principles. p.13.
[6] https://creativecommons.org/licenses/

Open Data means data that can be freely used, re-used and redistributed by anyone - subject only to the requirement to attribute authorship and share alike.

Researcher means a person who has entered into an employment relationship with UCT in an academic position, full-time or part-time, and whether full appointment or joint appointment, including honorary and affiliate appointments and assistantships.

4. Objectives of the Policy: Benefits of Data Availability and Reuse

4.1 Data as an Asset of Research

Data are important to the research lifecycle in all disciplines, both as a raw material and as a reusable output of research. Researchers have a fundamental responsibility to communicate findings and outputs, including the underlying data, to achieve the societal and economic development benefits of research.

4.2 Reproducible Science

Reproducibility is a core principle of research integrity, whereby an entire experiment or study can be duplicated by an independent researcher. Reproducibility is critical to the validity and progress of science and for assuring public trust in the research enterprise.

4.3 Benefits of Data Availability and Reuse

Open Data allows the validation of research results and reduced duplication of effort, which leads to more efficient research practice.

Increased international collaboration utilising readily accessible data, and their reuse in inter and transdisciplinary research, support a spirit of curiosity and accelerated innovation.

Reuse of data by citizens, private and public organisations leads to greater transparency and social engagement.

4.4 Open Access to Publically-funded Research Assets

Publically-funded research data are a public good, produced in the public interest and should be openly available free of charge to encourage extensive reuse.

4.4.1 Open by Default

In terms of the Open Access policy, unnecessary restrictions on the availability of research data should be minimised.

4.5 Principles for Reuse

The following principles for reuse are defined by the Royal Society:

"Data must be accessible and readily located; they must be intelligible to those who wish to scrutinise them; data must be accessible so that judgements can be made about their reliability and the competence of those who created them; and they must be usable by others."[7]

4.6 Legitimate Limitations on Availability

Necessary constraints on the availability of data include the protection of personal data; the protection of intellectual property; the protection of commercial interests of project partners; and security concerns.

4.7 Acceptable Limits of Openness

While the protection of personal data, commercial interests and security concerned are universally recognised as necessary limits on data availability, the reluctance of data creators to share data assets before these have been fully exploited is recognised. Ongoing research should not be compromised by premature data release.

4.7.1 Period of Exclusive Use

A privileged period of exclusive use is recognised in line with acceptable limits of openness, and should not exceed a maximum period of two years following the date of publication. Alternatively, the data may be released immediately but with an embargo. A user requesting access to embargoed data must enter into a data use agreement with the data provider. The embargo period should not apply to those data that support published findings and which are necessary for validation.

5. Scope of the Policy

The nature of data within the scope of this policy is twofold:

a. The data, including associated metadata that directly support or substantiate published research findings. Such data are required for validation, and should be made available concurrently with the research publication.
b. Valuable data assets that are created by the research project, but which may not directly underpin published research findings. Such data should be made available within the time

[7] Royal Society. Science as an Open Enterprise. https://royalsociety.org/~/media/Royal_Society.../2012-06-20-SAOE.pdf. p.7.12.14-15.

period indicated by normal practice in the discipline, but the embargo period should not exceed two years from the date of project completion.

It is recognised that not all data generated in a research project are worth sharing. A data management plan should be included in the project proposal, to describe the most significant and potentially valuable data assets to be generated. This should not preclude the sharing and retention of unanticipated but valuable data assets.

The software and tools required to generate and to replicate the data should also be made available. This includes the software, code, algorithms, protocols, analytical and visualisation tools. These additional resources are essential components of the data deposit.

B. PROCEDURES FOR IMPLEMENTATION

6. Criteria for Selection of Research Data

The characteristics of data that should be retained are varied, based on integrity, originality and contribution to the scientific knowledge base.

General guidelines for selection of data to be retained include:

- The data that substantiate published research findings.
- Significant data generated by the project.
- Unrepeatable observations.
- Longitudinal studies of human or natural events.
- Experimental results that would be impossible or expensive to reproduce.

The sharing of research data is expected where data constitute "unique resources", e.g. arising from high volume experimentation; low throughput data arising from long time series or cumulative approaches; and models generated using systems approaches.

Good practice in data selection is represented in the NERC Data Value Checklist.[8]

7. Stakeholder roles and responsibilities

[8] NERC *Data Value Checklist*. (http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist/)

This policy aims to support grant holders who collect, produce and re-use data, by defining the roles and responsibilities of stakeholders, including funders; researchers; the university as the research performing institution; and its data service providers.

7.1 Funder responsibilities

- Develop and communicate research data management policy.
- Define preference for non-proprietary international and community standards that facilitate access, use and interpretation of data.
- Specify retention periods for significant data.
- Endorse data deposit in existing domain-specific data infrastructure or other trustworthy repository.
- Specify required contextual information and metadata, including provenance, quality and uncertainty indicators.
- Provide advice directly or through data services.
- Review implementation of data management plans.

7.2 Researcher responsibilities

- Conform to funder policy in grant proposals, during project lifecycle, and on completion of the project.
- Define significant data to be produced on the research project.
- Document data storage and management requirements in data management plan.
- Submit completed data management plan for submission to funder with project proposal.
- Retain data during project lifecycle and for period specified by funder.
- Deposit data timeously with a University's repository or domain-specific data infrastructures or other trustworthy repository, according to funder policy.

7.3 University responsibilities

- Constitute a governance committee with oversight of policy implementation. (Annexure A)
- Ensure that compliance with policy requirements by grant holders is adequately supported.
- Define and maintain research data management policy at institutional level.
- Enable data management planning and execution of good data management practice by specific research projects.
- Raise awareness of researchers of policies, guidelines and associated data management practice.
- Support compliance with funder policies.

- Manage the data in the form in which it was originally collected, before summary, analysis and reporting.
- Ensure that the data is discoverable by means of appropriate persistent, unique and resolvable identifiers.
- Provide long-terms stewardship for some data, depending on national data infrastructure provision and eligibility.

7.4 Data service provider responsibilities

- Provide long-term stewardship for specific data, in accordance with funder policy.
- Provide guidance and support according to funder designated role.

8. Provision of Research Data Management Infrastructure

In the absence of an appropriate national or international data centre, including domain-specific infrastructures and those provided by the funding agency, the university is responsible for the long-term stewardship of the data. Research data management is integral to the research activity, and it is deemed legitimate to use public funds to support research data management effort. The cost of in-project research data management should be included in the grant proposal.

The cost for long-term preservation and access should be covered in a deposit charge paid from the research grant, provided this is transferred within the lifetime of the project, and provided that the designated repository does not also receive direct funding for this purpose.

The use of trusted or accredited repositories to provide long-term stewardship is recommended. Where these do not exist, institutional or other infrastructure should be used.

9. Data Management Planning

Researchers in receipt of grant funding may be required to manage their data; all researchers are encouraged to do so in accordance with good practice. Grant holders and their research groups are obliged to declare their practice related to data in a data management plan (DMP) to be submitted for assessment with the grant application. The University takes responsibility for quality assurance and monitoring the execution of the data management plan.

10. Discovery and Reuse

The value of data is realised in reuse. Shared data should be made easy discoverable, accessible and where possible, interoperable to specific international standards. A distinction is made between data that substantiate published research findings, and the data assets that are created by the research project. The former should be made available concurrently with the research publication, with a link to the web-based location and a persistent identifier. Where necessary, a statement should be included in the publication setting out reason why the supporting data cannot be made accessible.

Sufficient metadata should be recorded and made openly available to render the research data discoverable for effective reuse. The metadata standard applied should be appropriate, rich, of high-quality and comply with norms and good practice of the discipline by using a recognised standard where such exist.

A data citation is required to provide due recognition of the data provider, and should include "a persistent method for identification that is machine actionable, globally unique, and widely used by a community."[9] Digital Object Identifiers (DOIs) administered by the DataCite organisation are an example of such means to attach a permanent identifier to data.

Researchers are required to make data available using the least restrictive license possible. The use of Creative Commons license is widely recommended to ensure appropriate reuse, whilst still enabling Open Science.[10]

11. Recognition and reward for data providers

Data is recognised as a legitimate citable product of research, which requires the acknowledgment of data providers in data reuse and the citation of data that underpins further research findings.

To ensure appropriate recognition for collecting and analysing data, data providers are entitled to a limited period of privileged use, not to exceed a maximum period of two years, to enable the publication of research results.

12. Monitoring and reporting requirements

[9] Joint Declaration of Data Citation Principles: https://www.force11.org/datacitation

[10] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020p.11: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020[hi[oa[pilot[guide_en.pdf

Reporting against the data management plan (DMP) should be included as part of the annual and final report of the research project. Compliance in data sharing activities may be considered as part of the selection process for future funding proposals.

13. Effective Date

(a) Date of approval: 17 March 2018

(b) Approved by: University Research Council, Senate

(c) Date of implementation: RDM Implementation Plan, April 2018

14. Review Date

a) Date of next revision: March 2021

15. Related Policies

Intellectual Property Policy (2011)

Open Access Policy (2014)

Heritage and Collections Stewardship Policy (in draft)

ANNEXURE A

University of Cape Town

RESEARCH DATA MANAGEMENT GOVERNANCE COMMITTEE

TERMS OF REFERENCE

Status:

A URC sub-committee to provide oversight of research data management activities to ensure that compliance with policy requirements by grant holders is adequately supported.

Purpose:

The Research Data Management Committee (RDMC) is constituted to protect the interests of the University, researchers, funding agencies and the community to ensure consistent research practice related to data management principles that support effective data sharing, including open access; and the need for data to be discoverable, accessible, reusable and interoperable to specific quality standards.

Terms of Reference:

Consistent with its commitment to the responsible conduct of research, the University of Cape Town (UCT) is committed to ensuring compliance with policy requirements of grant holders in the implementation of the Research Data Management Policy.

In terms of the definition of research data (3.1), the RDMC will:

(a) provide oversight functions of funder mandates to preserve and openly disseminate data that support published research findings, and other data sharing imperatives;

(b) perform review and monitoring of grant-funded research projects where researchers may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research;
(c) approve exceptions to the public release of research data, for legal, ethical and commercial reasons;
(d) monitor associated policies, guidelines and practices to ensure that the research process is not damaged by inappropriate release of data;
(e) make available to Faculty Deans and to the URC, its activities and determinations (approval, approval with modifications required, non-compliance, etc.), to support consistent research data management throughout the University;
(f) monitor and evaluate institutional data management support activities provided to UCT researchers;
(g) communicate with stakeholders regarding awareness of policies, guidelines and associated data management practice;
(h) define and revise the Research Data Management Policy as required.

Meeting Procedures:

The committee will meet as necessary as determined by the Chair with a minimum schedule of twice per year, and on ad hoc basis for project review purposes; and to ensure timely continuing review (at least annually) of policy development.

Members will be requested to sign an attendance register and make apologies in advance, if unable to attend. Apologies should be made to the URC Servicing Officer based in the Research Office.

A formal agenda will be distributed electronically to all members, along with copies of all relevant material, two weeks prior to the meeting.

Minutes documenting main decision points, will be recorded.

A Principal Investigator (PI), who is also a member of the RDMC may answer any specific queries that other members wish to address, but must recuse him/herself prior to discussion and decision-making of the project.

Any PI may attend a meeting if requested to do so by the chairperson. PI attendance can be particularly useful if the project is novel or especially complex and where the RDMC would benefit from a full description of the data management activities.

Composition:

A designated representative of each Faculty Board and at least five members with relevant experience and expertise in the management of research data. The following persons are required:

a) an expert in the field of research administration;

(b) an expert in the field of library management;

(c) an expert in the field of eResearch;

(d) a designated institutional official responsible for research data management;

(e) an expert in the policies, applicable laws, standards of professional conduct and practice associated with research integrity;

(f) an expert in the field of research ethics.

Chair/Deputy Chair:

Chair: The nominee of the Deputy Vice-Chancellor (Research & internationalisation)

Deputy Chair: The Committee elects a Deputy Chair from its members

Special Quorum Rules

A quorum of 50% plus one must be present for the meeting to proceed.

Terms of Office:

Members are appointed for a period of three years, with appointment to consecutive terms if required.

Administrative Support/Servicing Officer of the RDMC

The RDMC Servicing Officer operates within the Research Office of the University of Cape Town in Mowbray, Cape Town.