

## RESEARCH DATA MANAGEMENT POLICY

<b>Document name</b>	Research Data Management Policy
<b>Owner</b>	Council
<b>Portfolio manager/Initiator</b>	Vice Principal: Research, Postgraduate Studies, Innovation and Commercialisation
<b>Approved by</b>	Council
<b>Date approved</b>	05 August 2021
<b>Date amended/revised</b>	

### 1. INTRODUCTION

The drivers and principles for managing research data at the University of South Africa are defined in response to the increasing number of policies published by funders of research to ensure the validation of research results. The funder requirements for research outputs are:

- To make provision for data reuse,
- To enable actionable and socially-beneficial science from publicly funded research.

Further, making data available advances scientific progress and innovation, and the consequent socio-economic benefit to society. Data sharing enables validation, replication, re-analysis, new analysis, reinterpretation or inclusion into meta-analyses.

The management of research data should be compatible with the University’s commitment to the highest ethical standards in research, protecting the rights, dignity, health, safety and privacy of the community, including research subjects, and with its commitment to the welfare of animals and the integrity of the environment.

UNISA retains the rights and ownership of research data from publicly funded research. Privately funded research requires an agreement to retain access or deposit rights in the UNISA data repository (Archive).

### 2. SCOPE

The policy covers UNISA employees and students research outputs, research data (paper-

based/electronic/digital format<sup>1</sup>) and subject to the provisions of any relevant contracts, funding or collaboration agreements.

### **3. PURPOSE**

The purpose is to ensure that research data is stored, preserved, retained, made accessible for use and reuse, and/or disposed of, according to legal, statutory, ethical and funding bodies' requirements.

This policy seeks to ensure consistent research practice related to data management principles that support effective data sharing, open access, and for data to be discoverable, accessible, reusable and interoperable according to quality standards.

### **4. DEFINITIONS**

Data set	is a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen;
Data management plan (DMP)	is a formal document that outlines how data will be handled both during the research project and after the project is completed;
Digital Object Identifier (DOI)	is an identifier (not a location) of an entity on digital works. It provides for persistent and actionable identification and interoperable exchange of managed information on digital networks. A DOI name can be assigned to any entity-physical, digital or abstract-primary for sharing with interested user community as intellectual property. The DOI system is designed for interoperability; that is to use, or work with existing identifier and metadata schema.
Format	is a specific manner in which information is encrypted and stored on a computer disk;
Metadata	is the structured descriptive information used to describe research data and associated documentation's attributes in a standard format. This is used for the discovery and retrieval of data, and can be used for citation and referencing;
Researcher	is the person(s) who originated or gave existence to the research, including UNISA employees and students;

---

<sup>1</sup> Please see Addendum 2 for format specifications

Research Directorate	is the Research Management Directorate of the University of South Africa, tasked with the operational implementation of the research and innovation strategy at UNISA;
Data repository	is an online archive for collecting, preserving and disseminating digital objects of the research data of UNISA;
Research data	refers to factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings;
Open data	is research data that is freely available permitting any user to download, copy, analyse, re-process, pass to software or use without financial, legal or technical barriers other than the requirement for attribution and share alike <sup>2</sup>
Open science	is the movement to make scientific research and data accessible to all, for public good. It includes practices such as publishing open scientific research, campaigning for open access and generally making it easier to publish and communicate scientific knowledge. Additionally, it includes other ways to make science more transparent and accessible during the research process. This includes open notebook science, citizen science, and aspects of open source software and crowdfunded research projects <sup>3</sup> .

## 5. RESPONSIBILITIES

Research data management (RDM) is a shared responsibility. The University assumes a coordinated approach to the administration and management of research data, through cooperation with different stakeholders. Researchers, academic departments and administrative departments should work in partnership to implement best practice of research data management.

### 5.1 Researcher

A researcher must submit a quality assured version of the research data into a repository. The principal researcher takes responsibility although not the person who will necessarily submit the research.

### 5.2 Information and Communication Technology Department

- a) Provide a data repository platform and other infrastructure for research data management for the submission of research data by researchers.
- b) Ensure the long term preservation of digital objects and continued access and use by future users.

---

<sup>2</sup> <https://sparcopen.org/open-data/>

<sup>3</sup> <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement/>

### **5.3 Funder**

The ownership or access rights will depend on whether the research is publicly funded by a government entity for the contribution to knowledge. Privately funded research should comply with UNISA's Research and Innovation Policy or Intellectual Property Policy.

### **5.4 College/supervisor**

Ensure the quality of the research data that is submitted for curation on the data repository.

### **5.5 Department: Research Administration**

Implementation of funder requirements for data curation and RDM.

### **5.6 Library**

Ensure curation of data, assign descriptive metadata and digital preservation.

## **6. DATA MANAGEMENT PLAN**

- 6.1 Researchers in receipt of grant funding may be required to develop a data management plan. All researchers are encouraged to do so in accordance with good practice.
- 6.2 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.
- 6.3 The University (the Research Administration Department) takes responsibility for quality assurance and monitoring the execution of the data management plan.
- 6.4 It is acknowledged that some data generated are more sensitive than others. Before initiating the research, it is the researcher's responsibility to consider the following:
  - a) confidentiality,
  - b) ethics,
  - c) security and
  - d) copyright.
- 6.5 Possible data sharing challenges should be considered in the data management plan with solutions to optimise data sharing.
- 6.6 The data management plan should indicate which data will be shared. If (some) research data is to be restricted, an appropriate statement in the data management plan and subsequent publication should explain why access to data is restricted.
- 6.7 Reporting against the data management plan (DMP) should be included as part of the annual and final report of the research project.
- 6.8 Compliance in data sharing activities may be considered as part of the selection process for future funding proposals.

## **7. DATA SETS AND PREPARATION**

- 7.1 Research data supporting the findings of UNISA research output and publications is curated and managed for data sharing and preservation.
- 7.2 Data must be prepared (e.g. anonymization) for deposition into the data repository, and submitted with the provenance and contextual metadata.
- 7.3 Depositing the research data in the data repository for any user to access, mine, exploit, reproduce and disseminate.
- 7.4 The data needed to validate the results presented in scientific publications, data which is valuable to others. Data which cannot be re-generated, must be submitted to the data repository.
- 7.5 Data underpinning a Masters dissertation and doctoral thesis: Study supervisors must be consulted/ensured of the quality of the datasets before any data is released. This is encouraged to ensure that the research process is not negatively influenced by premature and/or inappropriate release of data whereby one might be unable to use the data in future publications.

## **8. METADATA**

Assigning metadata to the dataset, including the funder award number and other attributes such as a *permanent identifier*<sup>4</sup> (Digital Object Identifier to identify one dataset uniquely from the other) for future citation and referencing as required by the publisher or funders.

Metadata should be made available irrespective of constraints that may exist in respect of data access.

Metadata records of both the data and the publication/Electronic Thesis and Dissertation (ETD) in UNISA repositories (UnisaIR and Data Repository) will include a link to the other where it is possible.

## **9. REPOSITORY**

The data repositories must comply with trusted repository criteria. The use of trusted or accredited repositories to provide long-term stewardship is required. Where these do not exist, institutional or other infrastructure should be used.

## **10. DATA SHARING**

Researchers share all research data that can be shared, and provide a link to the data in the publication.

Potential collaborators should be informed that due to public funding and funder mandate, the researcher is expected to share research data as openly as possible.

---

<sup>4</sup> Hudson, Simon, Molloy, Laura (2014). Report on *Current Best Practice for Research Data Management Policies* commissioned from CODATA by the Danish e-Infrastructure Cooperation and the Danish Digital Library. Version 2– 11.07.2018

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, after submission to the data repository.

## **11. RESTRICTIONS/EMBARGOES**

Publicly funded research data is in the public domain, with free and open access, by default. There are, however, a number of reasons for withholding research data, even if there are obligations to funders or publishers to openly share the outputs of research.

- 11.1 An appropriate embargo is acceptable in respect of an opportunity to publish. If a researcher needs to apply an embargo, reasons must be provided on why the data should not be in an open access repository.
- 11.2 Restrictions are acceptable if necessary to protect IP or commercially confidential data.
- 11.3 Personal/sensitive data should not be released into public domain, unless the person provides his/her written consent or the data is properly anonymized.
- 11.4 Data that can harm the interests of a community – human or otherwise – can be motivated as restricted.
- 11.5 Where data deposit is not possible or cost effective, it is acceptable not to publish the data in a trusted repository. However, the data repository must contain a metadata record that describes the data, the reason why the data is not available, as well as the alternative access to the data. The ability to validate the published research findings must not be compromised and the alternative access to the data should be provided.

## **12. 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 the 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.

## **13. RETENTION AND STORAGE**

To optimise research outcomes, data must be stored, retained, documented and/or described, made accessible for use and reuse, and/or disposed of according to legal, statutory, ethical and funding bodies' requirements.

- 13.1 Research data must be retained and guided by to the periods specified in the National Archives of South Africa Act 43 of 1996.
- 13.2 The specified period can vary depending upon the discipline and type of research.
- 13.3 Research data and materials must be stored securely to protect against theft, misuse, damage or loss. Research data must be held in appropriate facilities that allow access to be managed as required.
- 13.4 Research data must be retained for a period of 10 years, after which it will be appraise for further retention.
- 13.5 Once a dataset has been assigned by the designated South African Digital Object Identifier Mint Agent as a Digital Object Identifier, the data must be preserved in perpetuity.

#### **14. RELATED POLICIES**

- 14.1 Library Collection Development Policy
- 14.2 Policy for Master's and Doctoral Degrees
- 14.3 Data Privacy Policy
- 14.4 Information Security Policy
- 14.5 Intellectual Property Policy
- 14.6 Records Management Policy
- 14.7 Research and Innovation Policy
- 14.8 Research Ethics Policy
- 14.9 Policy for Copyright Infringement and Plagiarism
- 14.10 Policy on Grants from the Research Funds
- 14.11 UNISA Electronic Data Backup Policy

#### **15. COMPLIANCE**

- 15.1 The RDM Policy is a mechanism to aid UNISA researchers to comply with research funder requirements for RDM and data sharing.<sup>5</sup>
- 15.2 Joint Declaration of Data Citation Principles: <https://www.force11.org/datacitation>
- 15.3 The Data Citation Principles cover the purpose, function and attributes of citations. These principles recognize the dual necessity of creating citation practices that are both human understandable and machine-actionable.

These citation principles are not comprehensive recommendations for data stewardship. And, as practices vary across communities and technologies will evolve over time, we do not include recommendations for specific implementations, but encourage communities to develop practices and tools that embody these principles.

---

<sup>5</sup> <https://www.icsu.org/current/news/open-data-in-a-big-data-world-agreement-passes-120-endorsements>

The principles are grouped so as to facilitate understanding, rather than according to any perceived criteria of importance.

a) Importance

Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications<sup>6</sup>.

b) Credit and attribution

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

c) Evidence

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

d) Unique identification

A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.

e) Access

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

f) Persistence

Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.

g) Specificity and verifiability

Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.

g) Interoperability and flexibility

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities.

## 16. IMPLEMENTATION OF POLICY

---

<sup>6</sup> <https://www.icsu.org/current/news/open-data-in-a-big-data-world-accord-passes-120-endorsements>

The policy or any regulatory provisions which were in force prior to the commencement of this policy are replaced with effect from the date on which the Council approved this policy.

## **17. REVIEW OF POLICY**

This policy is reviewed every three years by relevant stakeholders under the custodianship of relevant committees.

## **ADDENDUM 1**

### **1. DATA MANAGEMENT PLAN**

- 1.1 This proposed plan is being presented to facilitate discussions on a design for a data curation process.
- 1.2 It is a framework for the recognition and development of the necessary tools and procedures to best serve the users.
- 1.3 Preserved data is more relevant, because data can be re-used by other researchers.
- 1.4 The researcher can direct requests for data from the database, rather than to acquire data individually.
- 1.5 Data that are preserved have the potential to lead to new, unanticipated discoveries and they prevent duplication of scientific studies that have already been conducted.
- 1.6 Data archiving also provides assurance against loss of data by the researcher.

### **2. MAJOR COMPONENTS OF THE DATA MANAGEMENT PLAN**

#### **2.1 Short-term management of data**

Version control for files, backing up data and data products, security and protection of data is important.

#### **2.2 Description of data to be archived**

The expected data to be archived includes:

- a) textual materials;
- b) databases;
- c) datasets;
- d) spread-sheets;
- e) software e.g. programming languages, operating systems, etc.;
- f) curriculum materials;
- g) digital images;
- h) audio and video formats;
- i) physical collections;
- j) models;
- k) experimental and observational data.

### **2.3 Acquiring of data**

Methods of how data will be acquired will be established.

### **2.4 Processing of data and system description**

- a) The specific software to be used, algorithms and the scientific workflow will be determined.
- b) Data are stored in a consistent format with related meta-information in a relational database.
- c) The data exchange between project partners and the availability of products and published data to the scientific community will be established through a client/server system and web services on the internet.
- d) Data are provided in various standard formats for harvesting by portals, catalogues and search engines.

### **2.5 Metadata content and format**

- a) Metadata are the contextual details, including and information important for using data.
- b) Metadata include descriptions of temporal and spatial details, instruments, parameters, units, files, etc.
- c) Standard metadata commonly used in the scientific discipline that contains the work and according to the standards of the chosen database will apply.

## 2.6 Quality assurance and quality control

- a) Any obligations that exist for sharing data collected will be addressed e.g. obligations from funding agencies, institutions, other professional organizations and legal requirements.
- b) Any ethical, privacy, copyright issues and also intellectual property will be addressed. To exchange unpublished data during the runtime of the project, data sets may be password protected.



## ADDENDUM 2

### 1. IMPORTANCE OF HOW DATA IS FORMATTED

- a) The format in which the data is saved will assist to maximize the share-ability and re-usability of research data.
- b) The careful selection of data format assists to limit the chances of data becoming obsolete when a proprietary format is no longer supported.

### 2. WHAT FORMAT SHOULD BE USED?

2.1 Formats which are more than likely to be accessible in the long-term are:

- [Open format](https://en.wikipedia.org/wiki/Open_format)<sup>7</sup> (or free file formats), non-proprietary
- Commonly used within your field or discipline
- Have a standard encoding (ASCII, Unicode)
- Not encrypted or compressed
- [UK Data Archive Recommended Formats](https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats)<sup>8</sup>

---

<sup>7</sup> [https://en.wikipedia.org/wiki/Open\\_format](https://en.wikipedia.org/wiki/Open_format)

<sup>8</sup> <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>

2.2 For further information please see the [list of preferred file formats](#)<sup>9</sup>.

- a) Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
- b) Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.



---

<sup>9</sup> <http://digital.lib.washington.edu/preferred-formats.html>