# POLICY:
# RESEARCH DATA MANAGEMENT

| | |
|---|---|
| Policy Owner | DVC Research and Internationalisation |
| Division/ Unit / Department | Postgraduate School |
| Date of Initial Approval | 18 March 2021 |
| Approved by | Senate |
| Last Date of Approval | 18 March 2021 |
| Approval date of Revisions/ Amendments | Reviewed 2021 |
| Next Review Date | 2026 |
| Platforms to be published on | UJ Website, Intranet |

# 1      PREAMBLE

The University of Johannesburg (UJ):

- In pursuit of its vision of being an international university of choice, anchored in Africa, dynamically shaping the future;
- In pursuit of its strategic goal of excellence in research and innovation as a component of its primary strategic goal to achieve global excellence and stature;
- mindful of its commitment to the sustained excellence and relevance of its comprehensive programmes and of its research;
- carrying out its mission of inspiring the community to transform and to serve humanity through innovation and the collaborative pursuit of knowledge, among others through research output;
- recognising its obligation to nurture employees and students with integrity, who are knowledgeable, well-balanced leaders and confident global citizens who are committed to the values which underlie an open and democratic society based on human dignity, equality and freedom (inclusive of academic freedom),

in this *Research Data Management Policy* states the rules and principles for the digital storage of and access to research data of UJ researchers.

# 2      RATIONALE FOR THE POLICY

This Research Data Management Policy of the University of Johannesburg (UJ) is a response to:

- the FAIR principles of Open Science[1];
- the requirements of funders to make data visible;
- a recognition of the importance of collaborative research;
- the need for greater research data integrity; and
- societal concerns that data should be more easily available via a digital platform.

The policy addresses both the principles of open access data and the need to limit the accessibility of data in certain cases, even when that data has been collected through publicly funded research programmes or projects.

---

[1] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M. & Bouwan, J (2016) The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data 3, 160018.

## 3     PURPOSE

The purpose of this policy is to:

3.1     ensure that all data collected by researchers affiliated to the University[2] is stored on a digital platform that is secure and easily retrievable;

3.2     ensure the integrity of data collected by making the data collection and storage process more transparent;

3.3     where appropriate, to make research more quickly and readily available on the University's digital Research Data Management (RDM) platform;

3.4     enable the cataloguing and identification of research data in order to increase the citation of this data by the research and educational communities;

3.5     encourage more collaborative practices amongst researchers through the sharing and re-use of research data; and

3.6     ensure that the long-term value of data is preserved by making the data readily accessible and usable in the future.

## 4     PRINCIPLES OF RESEARCH DATA MANGEMENT (RDM)

4.1     The RDM platform allows for open access to stored data or, if ethical, legal and/or commercial requirements are such that open access to data would negatively affect persons/institutions involved in a particular research project, for digital password-controlled access to the stored data, thus making it available to funders and, where appropriate, to other researchers.

4.2     In terms of the University Research Policy, the intellectual property of research data collected by researchers while in the employ of the University is held by the University.

4.3     Data, like a published research article, conference proceeding, book or chapter, is a scholarly product and should therefore be stored in a digital format that:

- is easily accessible;
- ensures that over time it will not be lost, corrupted or compromised;
- ensures that the data is given a reference identification code so that it can be cited by other researchers; and

---

[2] As defined in section 5 of the UJ Policy on Authorship.

- ensures that the data is stored in an ethical, legal and commercially appropriate manner according to the principles of the UJ Research Ethics Policy and the nature of the data;

- complies with the growing research data requirements imposed by funders and publishers; and

- enables researchers to save time and resources by managing their research data most effectively.

4.4 RDM facilitates the sharing of research data and, where appropriate, the re-use of data, in order to foster collaborative practices amongst scholars that maximizes the financial and intellectual benefit of research data collected and makes the most effective use of public funds.[3]

4.5 When research is published, the digital location of where the data can be digitally accessed upon which the research publication was based, should be provided in a footnote of any research publication that is not subject to an embargo.

4.6 To ensure that those who have collected the data are appropriately recognised for their work in collecting, storing, and analysing the data, all users of data must acknowledge through citation the researchers who generated, stored and shared the data.

4.7 Researchers must ensure that data is not released at any stage in the research cycle that is at variance with ethical, legal, and commercial requirements.

4.8 Research that is managed, stored, and shared does not include institutional data generated by UJ or the administrative data of a research project such as funders' policies and grant applications.


## 5    DEFINITIONS

The definitions used in this document are in line with the definitions provided by OECD *Principles and Guidelines[4]* for academic disciplines:

"***Research data***" means all factual records such as numerical scores, textual records, images, and sounds, used as primary sources in scientific research. This includes but

---

[3] University of Pittsburgh Research Data Management Policy. https://pitt.libguides.com/managedata

[4] OCED Principles p13.

is not limited to "notebooks, survey responses, software and code, measurements from laboratory or field equipment (such as IR spectra or hygrothermograph charts), images (such as photographs, films, scans, or autoradiograms), audio recordings, physical samples".[5]

"***Research data set***" means 'a systematic, partial representation of the subject being investigated'.[6] Specifically excluded are "laboratory notebooks, preliminary analysis, drafts of scientific papers, policies of funders, 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),"[7] which need not be stored on the RDM platform.

"***Data identification code***" means the digital object identifier (DOI) identification code that is generated for a data set when it is uploaded onto a digital RDM platform. Such a data identification code is globally unique and is a unique identifier assigned to every element of metadata and every concept/measurement of a dataset.[8]

"***Data Management Plan***" (DMP) means the plan established for the management of any data that is collected, stored, and shared.

"***Research Data Management***" (RDM) means the manner in which data is collected, stored, accessed and shared.

"***Data curation***" means the manner in which data is managed through the life-cycle of that data, including how the data can be used and re-used by other researchers in a scholarly or educational context.

"***Metadata***" means "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source".[9]

"***Open Access***" (OA) means "free, unrestricted online access to research outputs such

[5] University of Pittsburgh 'Research Data Management Policy'. https://pitt.libguides.com/managedata
[6] OCED Principles p13.
[7] OECD. Principles p13.
[8] FAIR https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/
[9] Riley, J. (2017) 'Understanding Metadata: What is Metadata, and What is it For?' Baltimore: National Information Standards Organization (NISO).
https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf.

as journal articles and books … [it] is open to all, with no access fees".[10]

"*Open Data*" (OD) means data that is free, unrestricted, and available online without any cost implications when this data is accessed either for view, re-use or redistribution.

"*Embargoed data (closed data)*" means "data to which access is restricted for legal, ethical, privacy and confidentiality and/or commercial purposes".[11]

"*Researcher*" means academic, academic support and research employees of UJ, researchers appointed by UJ in terms of Senate-approved policies and registered UJ students who were employed, appointed, or registered at the time the research was conducted and who are involved in the process of data gathering and/or analysis.

## 6.    OBJECTIVES OF THE POLICY

The objective of this policy is to ensure the safe, digital storage and accessibility of data and, where possible, to promote open access data by ensuring that data is stored in an accessible and re-usable manner as far as is ethically, legally and economically appropriate:

6.1    *Accessibility of data*: Research data must be accessible and easily locatable either in its raw form or once it has been analysed, coded or classified, so that judgements can be made about the reliability and validity of the data and the competency of the researcher(s) and research process through which the data was generated.[12] Research data generated through the use of public funds should be made available, as far as is ethically, legally and commercially possible, to the research and educational community for the public good.

6.2    *Data is an Asset*: In the context of research, data is an asset that needs to be properly managed to ensure that it can be made easily available and referred to in future.

---

[10] *Nature* (2019) https://www.nature.com/openresearch/about-open-access/what-is-open-access/
[11] University of Pretoria 'Research Data Management Policy'
https://www.up.ac.za/media/shared/12/ZP_Files/research-data-management-policy_august-2018.zp161094.pdf.
[12] This is in line with the Royal Society's policies on Open Enterprise. https://royalsociety.org/-/media/Royal_Soceity.../2012-06-20SAOE.pdf.p.7.12.14-15.

6.3 *Data should be re-useable*: A key objective in making data more easily available is to reduce duplication of effort and to thus increase efficiency in the research environment. It also offers an opportunity for researchers to build international networks with other colleagues and to accelerate innovation.

6.4 *Embargo limitations on the availability of data*: Not all data can be made equally accessible and legal, ethical, commercial and security requirements and concerns are considered before data is made openly available. Data that is not available through open access is stored on UJ's digital RDM platform, but access to the data is limited through digital password control.

5.5 *Open data by default*: Research that has been generated through public funds should be freely available to the public for the public good and to encourage the extensive use and re-use of data. As far as possible restrictions on the accessibility of data should be minimised.

6.5 *Period of exclusive use or embargo of data*: Where appropriate, a period of embargo or exclusive privilege can be set on the data to ensure that it is not open to the public. Should a user request access to data that is under embargo, the user must enter into a data use agreement with the data provider.

6.6 *Acknowledgement of data provider*: When data is stored on the RDM platform the data will be granted a unique data identification code which is used whenever the data is cited. This identification code is linked to the provider of the data who is then recognised for their work when the data or data set is cited.

## 7. SCOPE OF POLICY

7.1. The policy applies to all persons affiliated to UJ. These include:
- employees;
- postdoctoral research fellows (PDRFs);
- undergraduate and postgraduate students;
- research affiliates;

- visiting academics; and

- community engagement research collaborators.

7.2    This policy covers all data and metadata that directly supports or substantiates published research findings.

7.3    Valuable data which is collected but not used directly in a publication should also be stored and, where appropriate, made accessible to other researchers or to the public.

7.4    Not all data that is generated is worth storing or sharing.  The research data plan for each research project should clearly outline what data will be stored and shared and how this data will be stored and shared digitally.

7.5    Where appropriate, the tools required to generate and to replicate data should also be made available. This includes "the software, code, algorithms, protocols, analytical and visualisation tools".[13]


## 8.    CRITERIA FOR SELECTION OF RESEARCH DATA FOR DIGITAL STORAGE

The following guidelines[14] determine which data should be selected for storage:

- the data upon which published research has been done or will be done in the next two years;

- the data is drawn from unrepeatable observations;

- the study involved humans, animals, plants or natural events and is longitudinal in nature;

- the results of the experiment cannot be replicated because of the nature or cost of the experiment;

- legal reasons exist for storing the data;

- the data is of historical value;

- the data may have academic or educational value to others;

- the data has the potential for re-use; and

- the data addresses a contemporary or future problem.

---

[13] University of Cape Town (UCT) Data Management Policy.
[14] The list is derived from the NERC Data Value Checklist https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/ and the UCT Data Management Policy.

# 9. STAKEHOLDER ROLES AND RESPONSIBILITIES

The following are the roles and responsibilities for research data management of stakeholders involved in a research project:

## 9.1 Funder Responsibility

(i) Funders establish and communicate the RDM policy they require funded researchers to follow;

(ii) Specifies the period of time that data should be stored and retained; and

(iii) Gives advice to researchers regarding their DMP, where applicable.

## 9.2 Researcher Responsibility

9.2.1 The researcher develops a DMP that conforms to the funding requirements of their grant;

9.2.2 Implements their DMPs; and

9.2.3 Ensures that their data is stored timeously on UJ's digital RDM platform for the life-cycle of the research project and for the specified period required by funders, where applicable.

## 9.3 University Responsibility

(i) UJ establishes a RDM digital platform and makes it freely available to all UJ researchers. This platform facilitates open access to data, but also provides the embargo features for data that, because of ethical, legal, security or commercial requirements, does not have open access.

(ii) The UJ Library and Information Centre (LIC) implements, manages and maintains the RDM digital platform.

(iii) LIC advertises the RDM platform to all researchers.

(iv) LIC guides and trains researchers on the use of the UJ RDM platform in compliance with UJ policy and international best practice.

(v) UJ maintain a RDM policy and ensure that it is communicated to its research community.