

University of Cambridge

Research Data Management Policy

Framework

Last updated: 4 December 2019

The purpose of this policy framework is to provide guidance to research staff and students as to how they should manage and share the research data they produce in order to maximise the potential for research produced by the University of Cambridge in support of the [University's core values and missions](#). This policy framework is intended for researchers, research students and anyone who supports research at the University of Cambridge.

Background

1. The University of Cambridge is committed to disseminating its research and scholarship as widely as possible. In keeping with that commitment, it supports the principle that the results of its research should be freely accessible where possible and where appropriate, and therefore supports its staff and students in making their research data available 'as open as possible and as closed as necessary', as articulated in the University's [Open Research Position Statement](#).
2. University staff and students are responsible for managing and curating their research data in accordance with the [University Open Research Position Statement](#), the [policies of their research funders](#), the [University's Research Policies](#), the [University's Research Integrity and Ethics guidelines](#), the [University's Statement of Records Management Practice and Retention Schedule](#) and Chapter XIII of the [University's Statutes and Ordinances on Finance and Property, subsection Intellectual Property Rights](#). Key aspects of these documents are incorporated into this policy framework.
3. This policy framework will be reviewed annually by the [Open Research Steering Committee](#) and may therefore be subject to change.
4. The following terms used in this policy framework can be defined as follows:

Research data – “the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers [staff, students or those supporting research] in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others.”^[1]

Research staff and students - research staff are individuals undertaking academic research either in direct employment of the University or under formal agreement with the University in another capacity (e.g. retired staff) regardless of where that research is taking place, whether in Cambridge or elsewhere; research students are individuals who are undertaking academic research at the University in pursuit of a postgraduate degree or other award.

Active data – “Research data files that are in the process of continuous change and/or development. Files containing this data are accessed, amended and/or updated as new data is gathered and/or processed. Some datasets may never be ‘finished’. A ‘snapshot’ of active research data can be archived to create a version that is fixed and can be cited.”^[2]

Data Management Plan – a plan that outlines how data will be managed from the point of collection at the start of a research project all the way through to what will happen to the data once the project finishes. Typically a data management plan (DMP) will cover areas such as collection strategy, backup and storage of data, ethical/legal requirements related to data, data sharing and data archiving.

Metadata – information that describes significant aspects of a dataset. For example, this may include authors, title, date of publication, unique identifier, a description of what the dataset contains and licence. This provides other researchers with the information needed to understand and reuse the dataset as well as making the dataset more findable

Principles

5. Reproducibility of research is a central tenet of the University and all researchers are expected to conduct their work in a rigorous manner.
6. The University recognises the importance of the long-term availability, with as few restrictions as possible, and long-term preservation of all research outputs including research data.
7. As detailed in the [University’s Open Research Position Statement](#), the University is committed to implementing procedures that are discipline-appropriate, proportionate, evidence-based, practical, cost-effective and sustainable, and in the best interests of enhancing its mission, in order to disseminate research and scholarship as widely as possible from research publications to all forms of research data.
8. The University is committed to achieving compliance with the data policies of its external research sponsors, publishers and governmental agencies, and requires its research staff and students to abide by terms and conditions agreed with third parties. The University also recognises that such third parties’ policies are evolving and that they may require higher levels of data accessibility and dissemination in the future.
9. As detailed in the [Open Research Position Statement](#) the University recognises that “[a]cross the disciplinary spectrum there are a wide range of cultural settings that influence both capacity for and appropriateness of fully Open Research” and it

“encourages outputs of research, and where appropriate the accompanying data, to be ‘as open as possible, as closed as necessary’.”

10. The University recognises that there is a balance between openness and duties under professional codes and legal obligations.
11. This policy framework applies to all research conducted by University staff and research students.
12. The University supports the [Concordat on Open Data](#) and works in line with the [FAIR principles](#). It is working towards embedding these across the institution.

The University is responsible for:

13. Disseminating information amongst its academics about the requirements under this policy framework and under [policies of the University’s funders](#) in relation to research data. Schools, Departments and Faculties are expected to be proactive in disseminating these requirements within their communities and encouraging and facilitating compliance.
14. Developing and supporting infrastructure and services that enable good research data management to be practised across the institution. This includes, but is not limited to, infrastructure such as active data storage options, the University [Apollo repository](#), the research information system, and services such as the [Office of Scholarly Communication](#), [Cambridge University Libraries](#), [University Information Services](#), [Research Office](#) and [Cambridge Enterprise](#).
15. Providing training and guidance to promote best practice in data management and sharing amongst its research staff and students.
16. Managing a [website](#) providing guidance for the University’s academics in good data management practice, including the use of electronic lab notebooks.
17. Providing advice and guidance on issues connected with good data management, such as data protection, research integrity, research ethics and Intellectual Property rights.

University Research Staff and Students are responsible for:

PLANNING

18. Principal Investigator(s) should establish and maintain clear research data management responsibilities within their research group to ensure good data management is practised throughout the project and by all group members. In the case of collaborative projects, the Principal Investigators must jointly agree how data is managed and maintained, sometimes across different institutions.
19. Preparing a Data Management Plan (whether as a research group or individual), in accordance with guidance provided by the [University of Cambridge](#) and the [Digital Curation Centre](#) (DCC). If funders require a Data Management Plan, such plan needs

to be prepared according to [funders' requirements](#) and should also be updated at project close to record how data has been managed, archived and, where appropriate, shared. This is good practice for all projects.

20. Ensuring that legal, ethical and commercial constraints on release of research data are considered at the initiation of the research process and throughout both the research and data life cycles.
21. Allocating appropriate resources (time and financial resources) for data management in their grant proposal and throughout the duration of the project and at its close.
22. Updating their Data Management Plans throughout the duration of their project, and to ensure that at the end of the project all their research outputs, together with their location, are indicated in their Data Management Plans. These updates are a requirement of many funders.
23. Depositing their final Data Management Plans into an appropriate repository (discipline-based or [institutional](#)).
24. In collaborative projects, agreeing on IP and data ownership matters from the outset and ensuring these are made in alignment of institutional agreements, with legal advice where appropriate.

ACTIVE DATA MANAGEMENT

25. Abiding by any legal and ethical constraints pertaining to how data is stored, shared and retained.
26. Storing their data in an appropriate location, with at least two back-ups.
27. Ensuring the proper management of physical data/samples as well as digital data in line with local rules, funder and government requirements.
28. Ensuring that proper data citation guidelines are followed when reusing or referencing datasets.
29. Ensuring that wherever it is an eligible cost, appropriate funds must be fully costed onto research applications to meet all data management requirements, including the cost of archiving and sharing data.

SHARING AND ARCHIVING DATA

30. Making their research data underpinning published research findings as widely and openly available as possible, ideally by depositing research data in appropriate repositories (discipline-based or [institutional](#)). Such data should be assigned persistent Uniform Resource Identifiers (URIs), such as Digital Object Identifiers (DOIs) to increase its findability in accordance with the [FAIR principles](#).
31. Providing a statement in research articles describing how and on what terms any supporting research data may be accessed (or a statement that all data is contained within the article, if there is no supporting research data). Supporting data should be accessible online no later than the first online publication of the article.

32. Ensuring that published research data has appropriate metadata description in accordance with [guidance provided by the University of Cambridge](#).
33. Ensuring that research data records are retained in appropriate repositories (discipline-based or institutional) for as long as the data are valuable to the data creator or to others, or for as long as is required by the funder or by other regulatory requirements. If data is not publicly shared, researchers should ensure it is kept safely in line with funder retention policies and to ensure the data can be produced to demonstrate research integrity if needed.
34. When depositing research data into external data repositories, considering data repositories which support [Open Researcher and Contributor ID \(ORCID\)](#), funder IDs and grant IDs.
35. Storing publicly-funded research data that is not generated in a digital format in a manner to facilitate it being shared.
36. Upholding the University principles in rigorous, reproducible research that is persistently available by using, wherever possible, open source formats for files and file type [recommended for long term preservation](#).

Advice, guidance and training to meet the expectations set out in this framework are available through the University's [Research Data Service](#) and with the support of local [Research Data Champions](#) in faculties and departments.

References:

1. Definition taken from UK Concordat on Open Data
<https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-...>
2. Definition taken from Cambridge University Libraries Digital Preservation Policy
<https://doi.org/10.17863/CAM.32927>

Policy version	2
Date of policy review	23rd October 2019
Name of reviewer	Dr Lauren Cadwallader, Deputy Manager of Scholarly Communication (Research Data Management)

Original policy creation date	23rd April 2015
Name of creator	Open Access Project Board (OAPB), approved by Research Policy Committee. From June 2019, the 'creator' is the successor body to OAPB, the Open Research Steering Committee
Date for next review	23rd October 2020
Frequency of reviews	Annually
Review is the responsibility of	Head of Research Services and STEMM Libraries
Location of policy	https://doi.org/10.17863/CAM.10528