

The Chinese University of Hong Kong

Guidelines on Research Data Management

1. Introduction

The Chinese University of Hong Kong (CUHK) recognises research data as a valuable institutional asset to be managed. Good research data management enables research data to be securely stored, shared where appropriate, allows the verification of findings and supports digital preservation. In addition, the University acknowledges the role that good research data management can play in realising its vision to be recognised globally for the quality of its research, as well as achieving compliance with the data policies of external research sponsors, publishers and governmental agencies.

The Committee on Research Data Management (CRDM) was established under the Research Committee to propose guidelines on data management. The purpose of these generic guidelines is to provide guidance and support on the responsibilities of the University and all its research staff and students in managing, preserving, and sharing current or future research data. The CRDM will keep the Guidelines under review. More support can be found from the CUHK Library's Research Data Service.

2. Terminology

2.1 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

2.2 Research Data Management (RDM)

“concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results.”²

It covers creation of data and a plan for its use; the organisation, structure and naming of data; how to keep it secure, provide access when necessary, share active data with collaborators and more broadly publish and get cited, and archive when considered to be of long term value.

2.3 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.”³

2.4 Data Management Plan (DMP)

“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.”⁴

3. Guidance

3.1 Ownership of Data

In general, the following principles apply:

- 3.1.1 Where no external contract exists, the University will own the data generated from research by staff in the course of their employment at the University and within the scope of the duties described in their contract of employment. (See Paragraph 5.2(d) of Policy on Research and Section 5 of Policy on Intellectual Property)
- 3.1.2 The University allows the release of the data for non-commercial use provided that such release complies with the obligations and requirements. (see Paragraph 5.7 of Policy on Intellectual Property)
- 3.1.3 The University will own the data generated from research by research postgraduate students receiving financial support from the University and all other students (such as self-financed research postgraduate students, taught postgraduate students, undergraduate students, exchange students, visiting students in the course of research led by CUHK staff members) that involve use of the University’s facilities and resources. This ownership will not apply if the student enrollment is subject to a contract approved by the University (such as students enrolled in dual degree programme co-hosted by the University with another institute) that stipulated the ownership of the data created by such student. (see Paragraphs 6.1–6.3 of Policy on Intellectual Property)
- 3.1.4 Where research is carried out under a grant or contract as approved by the University, the terms of the agreement will determine ownership and rights to exploit the data.
- 3.1.5 In collaborative projects, Principal Investigators (PIs) must jointly agree on intellectual property, data ownership and RDM matters from the outset and ensure these are in alignment with institutional agreements and legal advice where appropriate.

3.2 General Principles

- 3.2.1 Allocate appropriate resources (time and financial resources) for data management in your project, throughout its duration and at its close.
- 3.2.2 Manage and curate research data in accordance with
 - CUHK’s Policy on Research, Policy on Intellectual Property, Policy in Protection of Personal Data (Privacy), University IT Policies, University Information Security

Policies, Honesty in Academic Work, and Guidelines on Student Assignment of Intellectual Property;

- Hong Kong legislation, for example the Personal Data (Privacy) Ordinance Cap. 486; and
- the policies, terms and conditions of your research funders

3.2.3 PIs are responsible for meeting the requirements on data management as specified in any agreement with funders.

3.2.4 When transferring cross-border data between jurisdictions, PIs are responsible for fulfilling the local regulations on cross-border data transfer. For data transfer between Mainland China and Hong Kong, the University's Guidelines on Cross-border Transfer of Research Data from Mainland China should be complied with.

3.2.5 Data generated from projects should be made as openly available as possible for sharing via the CUHK Data Repository or an appropriate subject-specific repository unless there are prior formal agreements with external collaborators and parties on non-disclosure or proprietary use of the data.

3.2.6 If the data generated from projects form part of a potentially patentable invention, the data creator must promptly notify ORKTS of such invention by submitting an invention disclosure form (See Paragraph 8.1 of Policy on Intellectual Property). Upon receipt of the invention disclosure, ORKTS will assist in formulating an appropriate patent strategy and handle the subsequent patent filing with CUHK as owner of the invention. The data related to the invention should be kept confidential until the patent application is filed.

3.2.7 The University encourages the adoption of FAIR data principles⁵ to make the data findable, accessible, interoperable and reusable:

To Be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2. metadata are accessible, even when the data are no longer available.

To Be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To Be Re-useable:

- R1. (meta)data have a plurality of accurate and relevant attributes.
 - R1.1 (meta)data are released with a clear and accessible data usage license.
 - R1.2 (meta)data are associated with their provenance.
 - R1.3 (meta)data meet domain-relevant community standards.
- 3.2.8 Ensure 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.

3.3 Planning

- 3.3.1 Establish and maintain clear research data management responsibilities within your research group to ensure good data management is practised throughout the project by all group members and in the event of their departure from the University.
- 3.3.2 Prepare and follow a DMP, whether as a research group or individual, and update throughout the duration of the project. If funders require a DMP, such a plan needs to be prepared according to funders' requirements.
- 3.3.3 Ensure that at the end of the project all the research data, together with their location, are indicated in the DMP to record how data has been managed, archived and, where appropriate, shared. In the DMP include clear information on collection, storage, the provision of long-term availability or destruction, accessibility, and conditions for data reuse. Finally deposit your final DMP into an appropriate repository (institutional or discipline-based).
- 3.3.4 Research postgraduate students and taught doctoral students who involve use of the University's facilities and resources are expected to comply with the general requirements on storing, managing, and sharing data as stipulated by the University and/or terms and conditions required by the funders under contractual obligation.

3.4 Data Acquisition and Analysis

- 3.4.1 Manage active data storage and security for the duration of the project. It is advised to enforce Second Factor Authentication (2FA) for creation, update or deletion of research data. The measure enhances the security of the project and align with the trends to minimize security risk for long term operation.
- 3.4.2 Ensure proper management of physical data/samples as well as digital data.
- 3.4.3 Abide with the research and publication ethics on the production, publication, and sharing of data. (See Paragraphs 5.4 and 5.7 of Policy of Research)

3.5 Data Dissemination and Preservation

- 3.5.1 Preserve your research data, and where possible publish it after the end of the project in line with your DMP. This may include depositing your data into the CUHK Data Repository or an appropriate subject-specific repository and assigning Digital Object Identifiers (DOIs) to increase its findability.

- 3.5.2 A suggested minimum retention period is five years after publication or after the completion of the project, whichever is later, and 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. PIs should ensure that a copy of the research data is retained by the project team after the completion of the research or in the event of their departure or retirement from the University.
- 3.5.3 Where possible it is expected that data will be made openly available through a minimally restrictive and voluntary common-use license. Data supporting a research output should be accessible online no later than the first publication of the research output. Researchers should ensure the data can be produced to demonstrate research integrity if needed.
- 3.5.4 There may be legitimate reasons why data cannot be made openly available. In these cases, where possible facilitated access to data for approved researchers is encouraged.
- 3.5.5 Data produced during human subject research must be anonymised as necessary to preserve privacy and confidentiality of research participants.
- 3.5.6 Store publicly-funded research data that is not generated in a digital format in a manner to facilitate it being shared.

4. Support for RDM

4.1 For support on research data service, please contact the Library's Research Support & Digital Initiatives team:

Website: <https://lib.cuhk.edu.hk/en/research/data>
Email: data@cuhk.edu.hk
Phone: (852) 3943 0821

4.2 For support on grant, contractual, knowledge transfer and intellectual property issues, please contact ORKTS.

4.3 For support on data security, please contact ITSC.

First version:

Endorsed by the CUHK Research Committee on 15 July 2021;

Approved by the CUHK Administrative and Planning Committee on 10 August 2021.

Revised version:

Endorsed by the CUHK Research Committee on 17 January 2024;

Approved by the CUHK Administrative and Planning Committee on 9 April 2024.

References

¹ Concordat on Open Data (2016). <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>.

² Whyte, A., Tedds, J. (2011). "Making the Case for Research Data Management." DCC Briefing Papers. Edinburgh: Digital Curation Centre. <https://www.dcc.ac.uk/guidance/briefing-papers/making-case-rdm>.

³ Langley, S. (2018). Cambridge University Libraries Digital Preservation Policy. <https://doi.org/10.17863/CAM.32927>.

⁴ University of Cambridge Research Data Management Policy Framework. <https://www.data.cam.ac.uk/university-policy>.

⁵ "FAIR Principles." GO-FAIR. <https://www.go-fair.org/fair-principles/>.