Proposed research data retention and disposal workflow: staff & student projects

Research project lifecycle

Planning stage / HDR Milestone 1

Research stage / HDR Milestone 2

Close out stage / HDR Milestone 3

DMP creation Collect metadata about the data, including description, governance, access & retention.

Provide access to R: drive

R: drive Frequent access tier

Other data storage options e.g. Dropbox, laptop

If no access within 30 days

R: drive Infrequent access tier

If no access within 90 days

R: drive Glacier instant retrieval tier

Send communication about data retention
- to HDR students on M3 completion
- to staff on financial close out

Researchers decision: data is inactive — No → No further data retention action required

Yes ↓

Researchers prepare inactive data for archiving, publishing, retention & disposal according to University guidelines.

Researchers revise their DMP. Collect appraisal, preservation & sharing metadata and trigger date.

DMP Tool workflow to alert CIMA of projects where trigger date & disposal date are not empty.

Retention period is greater than 7 years? — No

Yes ↓

CIMA approve metadata & retention period — No

Yes ↓

DTS move dataset to deep archive tier. Update project data status in DMP record.

Abbreviations & definitions

CIMA: Curtin Information Management and Archives
DMP: Data management plan
DTS: Digital & Technology Solutions (Curtin's IT service)
R: drive: Research data storage drive
W: drive: proposed deep archive tier. Read-only access to researchers
Trigger date: date data deemed inactive
Disposal date: date retention period expired

Research data retention

Retention

Appraisal

Destruction

DMP Tool system identifies project data disposal date has passed

W: drive Deep archive tier ← Extend retention — CIMA appraisal — Destruct → Destroy data

Metadata to collect during planning stage or HDR Milestone 1

Description:
Information on why, what, where, how & by whom the dataset was collected. This can include title, instrument, date range & geographic information. Information can be retrieved from other Curtin systems. Useful persistent identifiers are RAiD, ORCID.

Governance:
Information about ownership & stewardship of the data. Information can be retrieved from other Curtin systems. Useful persistent identifiers include RAiD, ORCID & ROR.

Access:
Information collected to ensure appropriate custodianship of the data. Includes embargoes, restrictions & access control.

Retention:
Information about the significance of research to determine appropriate retention period in accordance with the Western Australian University Sector Disposal Authority.

Metadata to collect during close out stage or HDR Milestone 3

Retention:
Additional information include trigger date (date data deemed inactive) & disposal date (end of retention period).

Appraisal:
Information that can assist the institution to make decisions on whether the dataset should be destroyed or kept beyond its retention period. Includes research outcomes & future application of research data.

Preservation:
Information that can assist the institution to enable long term preservation & access. Particularly important to datasets that have longer retention period.

Sharing:
Information to help ascertain if the dataset has been published to an institutional or third-party repository. Information can be retrieved from other data sources. Useful persistent identifiers include RAiD & DOI.

Recommended steps to prepare inactive data for archiving

- Review content in R drive:
  - Remove ephemeral and duplicated files and data.
  - Ensure that research data and code, as well as project and data documentation (for example: project plans, ethics documentation, data dictionary and 'read me' text files[1]) are kept in R drive.
  - Describe or include (subject to licence restrictions) the software required to open the data files.
- To ensure the dataset can be used and accessed for the long term, include data files in open file formats as well as in the original format, if possible. For further guidance, check the recommended and acceptable formats by the UK Data Service[2] and the Library of Congress[3].
- Assign a data manager, who can approve requests for access to the data, make decisions about the data and is accountable for policy and legislation compliance. Data manager is usually the HDR supervisor or Principal Investigator.
- Consider publishing the dataset to the Curtin Research Data Collection[4].
- Even if the research dataset has been published to an open data repository or Curtin Research Data Collection, an archive copy is still required to be stored in the R drive.

1 https://libguides.library.curtin.edu.au/c.php?g=202401&p=1333152#s-lg-box-21330017

2 https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/

3 https://www.loc.gov/preservation/resources/rfs/format-pref-summary.html

4 https://www.curtin.edu.au/library/about/research-data-collection/