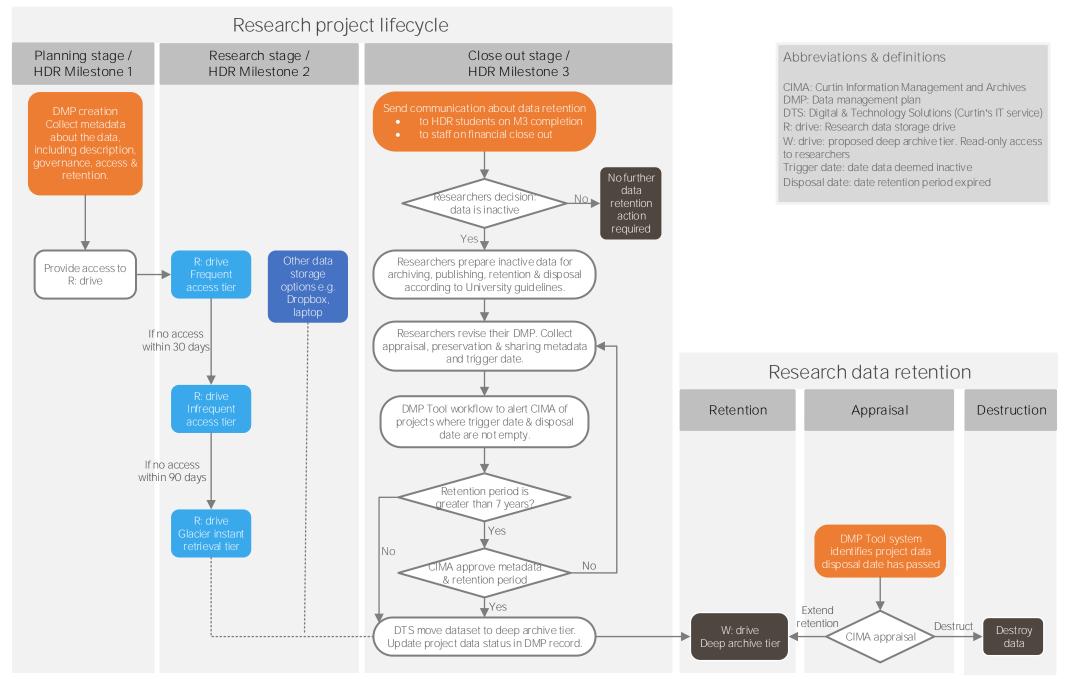
Proposed research data retention and disposal workflow: staff & student projects



August 2022. ARDC Institutional Underpinnings Program, Curtin University implementation project output. https://doi.org/10.25917/EDCN-5788

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:
 - o Remove ephemeral and duplicated files and data.
 - O 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 files¹) are kept in R drive.
 - O 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 Service² and the Library of Congress³.
- 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 Collection⁴.
- 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/