

# ARDC Institutional Underpinnings

## Element: Research Data Retention and Disposal

30/03/2022

### CONTENTS

EXECUTIVE SUMMARY	1
DESCRIPTION OF THE ELEMENT COMPONENT	2
DIFFERENCES IN APPROACH AND NEED	3
RECOMMENDATIONS AND ADVICE	4
WORKING GROUP ACKNOWLEDGEMENTS	5

### EXECUTIVE SUMMARY

Institutional Underpinnings is part of the ARDC's National Data Assets Initiative. In this program, 25 Australian universities are collaboratively developing a national Institutional Research Data Management (RDM) Framework. This Framework is intended to inform institutions' design of policy, procedures, infrastructure and services, and improve coordination of RDM within and between institutions. This output describes the initial findings of the research data management Retention and Disposal Element of the Framework, providing institutions with guidance to ensure the mandated retention and disposal of ever increasing volumes of research data on appropriate systems in line with legislative and funding body requirements. The output highlights the importance of embedding research data metadata and persistent identifiers to enable integration across research administration and management systems. Finally, the output recommends the development of a decision making model that provides triggers for

the appropriate retention and disposal of research data. Recommendations for institutions and Calls to action are highlighted throughout the Element. Calls to action specifically identify the need for future collective action from institutions and the community. This initial research data management framework Research Data Retention and Disposal Element will be further developed through additional institutional consultation and will be complemented by activities to validate and test the outputs described within.

## DESCRIPTION OF THE ELEMENT COMPONENT

Research data is an immense, diverse and valuable resource that requires specific management by researchers and institutions over extended periods of time. This is particularly evident once a research project's active phase has been completed, when long-term plans for data storage, retention and disposal need to be implemented, and the responsibility for on-going management of data is passed to different stakeholders and administrators. Storage, retention and disposal of research data require specific management actions that take place often many years after the research project has been completed, so processes need to be developed that ensure that the information needed to drive future management decisions and processes is carried forward. These processes are integral to reining in burgeoning storage costs while enabling a FAIR<sup>1</sup> approach to research data, as they ensure long-term management provides access to data for re-use in subsequent research.

### Storage

The rate of data generation in research is growing exponentially, presenting institutions with a 'grand challenge' to provide adequate and appropriate storage infrastructure at an affordable cost now and into the future. It is important to recognise that not all research data is of the same 'value' to institutions and researchers, and there may be substantial advantages in applying a data classification methodology to determine appropriate storage requirements. For example, research data can be moved to less costly 'low tier' storage options once a research project has been completed, with storage costs decreasing for less frequently accessed data.

It is apparent that mechanisms and processes need to be developed that allow institutions to evaluate the status and management requirements of stored data so it can be managed optimally, and that these mechanisms move data to appropriate long-term retention and curation platforms according to a predetermined timeline.

---

<sup>1</sup> <https://ardc.edu.au/resources/aboutdata/fair-data/>

## Retention

Management of data beyond the end of a research project's active phase is complex. Both researchers and their institutions are required to store research data on appropriate systems for mandated retention periods, specified through legislation and funding body requirements. Retaining accurate records of research processes, datasets and outputs is integral to defending the integrity of all research findings, allowing published research to be reproduced and verified.

Research data are required to be retained for a minimum of 5 years, extending to much longer periods (including permanent retention) so access to and management of long-term curation platforms is essential. Effective long-term management of data at scale can only be achieved by implementing automated processes that rely on adequate metadata and persistent identifiers being provided for each dataset.

## Disposal

After mandatory retention periods have elapsed, actions need to be implemented that consign data to various disposal processes (should the data be reused, destroyed, or retained indefinitely?). At predefined trigger points, data needs to move to an 'end state' (disposal) based on decisions/requirements that have been made by various stakeholders, to fulfill

1. obligations to funding bodies, commercial partners, publishers
2. expectations of research communities and reviewers
3. regulatory requirements

Typically, these actions occur many years after the research project has concluded. We need to ensure that all relevant data management information (plans, persistent identifiers, metadata) is carried forward with a project's research data so the appropriate disposal directions can be applied.

**Call to action 1:** Institutions are encouraged to share with one another examples on predetermined timelines and trigger points appropriate for managing retention and disposal

## DIFFERENCES IN APPROACH AND NEED

While all research institutions are required to comply with research data retention legislation and funders' requirements, most have adopted different approaches to managing data and investing in storage infrastructure, and show different stages of 'maturity' in implementation of research data management processes.

Funding of infrastructure (including provision of cloud storage) plays a fundamental role in management of research data. Generally, research-intensive institutions have larger amounts of data to store and manage, but also attract greater funding to help finance infrastructure, whereas smaller institutions need to provision suitable data storage and retention platforms within a tighter budgetary environment. Accordingly, each research institution has its own flavour of storage and retention platforms and have adopted different approaches to moving and managing data across systems. This lack of standardisation makes sharing and re-use of data difficult, which hinders research outcomes and also impedes movement of researchers and data between institutions.

We propose the adoption of a standardised approach to the storage, retention and disposal of data. The following recommendations are institution-agnostic and provide guidance towards the consistent management of research data within and across institutions.

## RECOMMENDATIONS AND ADVICE

The key to standardisation in data environments is to embed the use of metadata and persistent identifiers for datasets across systems and research projects. This enables development of connected systems and drives machine-actionable data movements throughout the research data lifecycle, but will be particularly valuable for long-term data management.

Each institution will need to adopt a decision-making model that will be used to define and implement the processes (automated and manual) that govern management of data within research projects. The processes and governance should determine the key metadata and identifiers (associated people and project context) that are required.

### **Recommendation 1:** Adopt data standards and identifiers that link researchers and research projects

Identifiers include e.g. Digital Object Identifiers (DOI) for publications and data collection, Open Researcher and Contributor IDs (ORCID), Research Activity Identifiers (RAID), Field of Research (FOR) codes, Research Organization Registry (ROR) codes etc. Use of data standards and standardised identifiers is beneficial to researchers and to administrators, and allows research projects, data and personnel to be managed across multiple systems.

### **Recommendation 2:** Design integration tools for administrative and research management systems

Internal integration enables communication and alignment of processes and systems which is vital to the development of automated data and research management functions e.g. grant administration, research metrics, publications, records management, research data storage. Note: use of common identifiers and metadata are critical to effective systems integration.

**Call to action 2:** Institutions are encouraged to share with one another examples of metadata and identifiers for system integration

**Recommendation 3:** Develop a broadly applicable decision-making tool (matrix, decision tree, rubric) that can accommodate local policy

A decision-making model will identify trigger points where actions and processes need to occur. Trigger points should be time-based and relate to various events being completed e.g. end of research project, publication of paper etc. The triggers will be defined by each institution, as they will drive development and implementation of local specific processes (e.g. define end of project process, use this as trigger event to automatically move data into inactive storage). Consequences of each trigger will vary according to institutional policy and legislative obligations, and any tool needs to accommodate local requirements (e.g. retention period legislation, archiving policy, available storage platforms).

Characterisation of use cases will inform local requirements for decision tool development. In many cases, well-validated use cases can be found in outliers, as they are small in number and known for their specific requirements e.g. large capacity storage uses, or require strict access control policies.

**Call to action 3:** Institutions are encouraged to make available to one another use cases that show approaches for retention and disposal decision-making.

**Recommendation 4:** Ensure contemporary capture and propagation of metadata for research projects

This promotes accuracy and utility, and enables appropriate decisions to be made regarding retention and disposal of data. Combined with unique researcher and project identifiers (recommendation 1), specific metadata should be added to all research data and as early as possible, through multiple touchpoints (e.g. grant application, human research ethics application, research clearance, resource provisioning). The metadata should be a single point of capture, and enable machine-actionable processes that drive auto-population across systems and minimise user burden (avoiding duplication of inputs). These metadata also facilitate movement of data through storage, retention and archiving platforms.

## WORKING GROUP ACKNOWLEDGEMENTS

Max Wilkinson (Chair)	ARDC
Adele Haythornthwaite (Editorial Committee)	University of Sydney
David Abramson	University of Queensland
Kylie Burgess	University of New England

Luke Burns	Edith Cowan University
Janice Chan	Curtin University
Racheal Chesters	University of Southern Queensland
Jacky Cho	University of New South Wales
Kate Cumming	University of Sydney
Kaye England	University of Southern Queensland
Sandra Ennor	Monash University
Erin Gallant	Australian National University
Owen Griffiths	Griffith University
Dianne Jeffery	Curtin University
David Jung	University of New South Wales
Jane King	Swinburne University
David Poger	Microscopy Australia
Peta Scott	University of New England
Kay Steel	Federation University
Chris Stueven	University of Melbourne
Jared Winton	University of Melbourne

The outputs of this working group were edited for public release by Frankie Stevens, Lyle Winton and Nichola Burton (ARDC)

Suggested citation: Australian Research Data Commons. (2022, April 1). ARDC Institutional Underpinnings Framework draft release. Zenodo. <https://doi.org/10.5281/zenodo.6392340>

Licence: [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)