Australian Research Data Commons

# ARDC Institutional Underpinnings Remaining Elements

30/03/2022

## CONTENTS


EXECUTIVE SUMMARY 1

CYBERSECURITY 1

DATA OWNERSHIP 1

DIGITAL PRESERVATION 2

FUNDING AND SUSTAINABILITY 3

GOVERNANCE 3

IDENTIFIERS AND METADATA 4

NON-DIGITAL MATERIAL 5

STANDARDS AND GUIDELINES 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. The following elements of the Framework were identified by participating universities in early workshops and feedback, but were not further developed by the expert working groups. While all elements were considered priority areas for universities, the expert working groups focused on an initial eight elements that were determined via potential benefit and likely phase two projects. Captured here are

explanations of the remaining elements, their importance to Australian research institutions, and the related challenges and opportunities. Also captured were key questions resulting from the participant discussions. Institutions may find these questions useful to consider when addressing these areas of RDM. These questions will also help to guide future collaborative work in the sector.

# CYBERSECURITY

Appropriate cybersecurity is the cornerstone for keeping research data safe and secure. It is an important aspect of good research data management (RDM). Many institutions have IT departments that have dedicated cybersecurity expertise to oversee their institution, but need input from research data specialists to understand and support the particular needs of research data. Researchers are required to ensure that their data has been stored securely, for the long term, and sometimes with appropriate access controls in place to comply with legal, ethical, and funder requirements. Institutions therefore need to provide guidance on the level of security attached to various institutional data services, and how to appropriately choose and correctly utilise them. A key consideration is providing researchers and institutions with an overview of cybersecurity elements.

How are we building cybersecurity into infrastructure in support of and on behalf of researchers? What cybersecurity actions are required and what are being applied by institutions to secure research data (e.g. encryption, enabling multi-factor authentication)? How do we monitor the security of and access to data? What constitutes a notifiable research data breach and how are institutions managing, identifying and being made aware of breaches? Is there a best practice model? What is the indicative list of (research) legislation relevant for cybersecurity? What are the institutional responsibilities including systems and provision of training, and who needs to be involved (eg. IT, ethics, research services)? What are the responsibilities of and requirements on researchers? How are researchers made aware of what is supported? What cybersecurity training is available at institutions?

# DATA OWNERSHIP

A national RDM framework could provide guidance on data ownership and usage rights, seeking harmonisation with institutional policies on data ownership with the aim of achieving a national approach.

How clearly is this stated in institutional policies at this point? Is clarity required on data ownership, retention rights, usage rights and publishing rights and limitations when working with complex projects, commercial entities, on industry-funded projects, with governments, with international partners? How do institutions detect and address gaps when Intellectual Property (IP) and RDM policies overlap or interact? How do we address citizen science project data ownership especially as researchers move across institutions? How might we better manage ownership and access when researchers and students

move institutions or move on? How do we manage ownership and access when data sources retain ownership, or impose usage restrictions, or include complex agreements from multiple providers? How do we better manage the audit trail of custodianship? How do we record, define and manage compliance with indigenous data ownership, community ownership and data sovereignty?

# DIGITAL PRESERVATION

Digital preservation (DP) involves the processes of assessing, collecting, organising, maintaining, preserving, and providing continued access to content determined to have long-term value, in ways that ensure authenticity, reliability, integrity, and usability [UNESCO and ICA '*Universal Declaration on Archives*', 2011[1]]. The goal of DP is to ensure that digital assets are managed well and cared for to remain usable and understandable over long periods of time - through both technology and organisational changes. This involves institutional elements such as policies, strategies, and institutional commitment to maintain and provide access to digital content.

Like other elements, institutional DP intersects with other RDM framework elements. But DP as an institutional element of an RDM framework goes beyond, providing cultural awareness and training in preservation techniques to academics. DP also goes beyond infrastructure practices to ensure the safety and accessibility of data, such as in long-term retention (e.g. bit-level preservation, including, but not limited to, checksum monitoring). DP includes a focus on content preservation and technological dependencies (software and hardware), which incorporates developing and maintaining risk profiles of digital collections in an institution's custody, technology watch processes, through to preservation actions (e.g. identification of existing file formats; migration of proprietary and/or legacy file formats to open and/or recommended preservation file formats). DP also includes establishing safeguards against accidental and/or malicious deletions or corruptions, agreed-upon actions for rectifying of issues discovered, and complex standardised metadata support (e.g. descriptive, technical, and preservation metadata to aid in providing context, ongoing maintenance and improved discoverability). While retention processes may partially facilitate ongoing access to digital assets, a more comprehensive institutional DP approach is necessary to ensure long-term reuse, integrity and reproducibility.

How do institutions select research data assets of significance and enduring value, while balancing with the best use of capacity? How can institutions select assets while taking into account their institution's unique profile, their custodial roles and key stakeholders? When does selection occur? How can identification be built into all stages of the data lifecycle? How are institutions committing to long-term research data preservation in terms of workforce, ongoing functions and infrastructure? How do we foster a research and RDM aware workforce skilled in digital preservation as the importance of this institutional responsibility increases? What funding models work for ongoing and long-term preservation of assets, given the different timelines and needs of active data management and IT

[1] https://www.ica.org/en/universal-declaration-archives

support. Can a more holistic perspective be achieved in the interests of maintaining assets across segmented approaches to resourcing?

# FUNDING AND SUSTAINABILITY

Advice is needed for research institutions, researchers and research support staff, on funding and seeking funding for RDM activities. Guidance may also be needed for funding agencies mandating RDM. Clarity is required on components and proportion of research funding and institutional resourcing that can be used for RDM activities. Specific guidance is required on how to fund RDM sustainably.

RDM should be planned and factored as part of research funding by researchers and institutions, to be sustainable and cost-effective. RDM will scale to varying sizes and with varying priorities of research institutions. What should funders mandate with regard to RDM policies or RDM Planning? What can researchers ask funders for in terms of funding for RDM services - rules and expectations? Resourcing models for the RDM services will differ, but what percentage should be expected to come from institutional block grants? How can institutions aggregate different sources of funding in the most efficient way, promoting recognition of the role of co-funding, and helping researchers maintain a sense of ownership? How can institutions best address people resourcing as well as infrastructure funding challenges? This should be viewed as a responsibility for the enterprise. Is it possible for institutions to support all RDM for all research activities? Are costs to smaller institutions disproportionately challenging and can larger collaborative use help? Are institutions being driven to engage external providers and at higher costs? What is the role of ARDC and National Research Infrastructure in funding and sustaining RDM? How can non-traditional research outputs (NTROs) be sustained, curated and managed given they provide further challenges?

# GOVERNANCE

There are several levels of governance to be defined and addressed by the framework. Governance cascades through several levels from institutional governance of the RDM process down to individual responsibilities for mediated access to restricted data. Levels can also include governance over multiple projects, programs of long running studies, or governance specific to projects. Individual institutions have responsibilities for aligning with the *Australian Code for the Responsible Conduct of Research* and its guidelines on RDM[2]. Repository governance is important, with many institutions managing institutional repositories, some hosting research data, and with some researchers using disciplinary repositories to both store and share data. Some datasets generated by research require ongoing and mediated access, eg. via an expression of interest process for usage and access that is assessed on merit

[2] "Australian Code for the Responsible Conduct of Research", 2018.
https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018

and as appropriate; Australian Bureau of Statistics restricted datasets; locally curated datasets; population health characteristics data only to be shared with registered users). Such a process can also require subsequent publications or resulting research be properly attributed and tracked. Furthermore, collaboration with other institutions can involve joint responsibility for shared data which may require multi-institutional governance.

What is optimal for institutional governance, what does it need to address, and what models exist? Is a separate governance structure required for RDM or is it covered by other research governance structures, such as in ethics? Do existing data governance structures include risk management assessment and planning around data security, privacy, IP and research reproducibility? Are there data and information management policies already in place that include research data and how are such policies enforced? How are governance structures communicated and how are academics best engaged?

What are the key roles for institutional governance of research data? Is there a need for defined multiple roles for example, data owner, data steward, data custodian, data manager? Are there examples or guidance on how to map such roles to researchers, research support and administrative staff? Does governance play a role in ensuring researchers' needs are adequately met with infrastructure and appropriate types of storage? How do we ensure the long tail of smaller data is adequately governed as well as big data?

How does governance address the transfer of custodianship when researchers leave, or the ongoing management of data once a project is complete? How are institutional repositories governed and what are the related policies around data deposit, data curation, data preservation, and data access? Could open research be a driver to kickstart discussions around inadequate areas of research data governance? Is mediation of research data access an inappropriate use of researchers' time, what models exist to manage this task, and how should institutions support this?

For data that has multi-institutional requirements, where does governance for this kind of data sit, how is this agreed (eg. prior agreement, contract), what effect do legal jurisdictions have, and how are issues resolved?

# IDENTIFIERS AND METADATA

The importance of a range of different identifiers should be highlighted. Institutions require guidance on best practice, metadata management and responsibility of managing metadata within institutions. The metadata goal is to enable data to be identified, retrieved and managed over time by capturing information including data contents, quality, format, collection/creation methodology, storage location, usage requirements, management requirements, and people involved with collection. The identifier goal is to unambiguously reference resources (digital and physical) within a given context, typically by means of a string or number conforming to a formal identification system (such as Dublin Core Metadata

Schema). Guidance is required on how to practically apply and manage the various identifiers in use, e.g. Digital Object Identifiers (DOI), Open Researcher and Contributor IDs (ORCiD), Research Activity Identifiers (RAiD), Research Organization Registry (ROR). This includes how to use identifiers to link systems, when to use local or public, persistent or temporary identifiers, and which identifiers to use for consistency between institutions.

With the proliferation of identifiers and their use, how do we minimise/standardise? What identifiers are required by funding bodies and how are they used? How are identifiers linked to impact and researcher incentives? What do we mean by metadata, what are the standards or vocabularies or data dictionaries, and what is discipline specific metadata? What are the responsibilities of researchers, institutions and support staff in capturing and maintaining metadata and identifiers? How can machine readable metadata and identifiers play a role, especially in making data "findable, accessible, interoperable and reusable" (FAIR[3])? How can metadata be practically applied to datasets across multiple platforms and research environments, especially when data is moved, copied or transferred?

# NON-DIGITAL MATERIAL

Guidance is required for institutions and research fields on the objectives and management of non-digital research material to align with the guidance related to digital data, records and other digital objects.

How do we advise and engage on planning for, management of, and digital integration of non-digital or non-digitisable primary material? How does this supplement, support or work together with guidance on digital materials? Is there a need for a national approach (eg. biobanking) or identifier strategy for non-digital and/or non-digitisable primary material? How should field-specific standards and identifiers relating to non-digital primary materials interact across fields of research. Should shared vocabularies be promoted relating to these primary materials? Should RDM policies encompass non-digital / non-digitisable primary material? Should digital and non-digital material be managed separately or integrated? How should the decentralised management and cross-field nature of non-digital materials be addressed?

# STANDARDS AND GUIDELINES

Standards and guidelines are both important to this framework but do they need to be addressed separately? Standards here are defined as comprising the mandatory legislation or other types of regulation with which institutions must comply, whereas guidelines are not necessarily enforceable.

[3] https://ardc.edu.au/resources/aboutdata/fair-data/

Standards include the *Australian Code for the Responsible Conduct of Research*[4], international standards such as General Data Protection Regulation (GDPR), and national and state legislation and regulations around data, archiving, retention and disposal. Guidelines may comprise local institutional recommendations, processes, 'how to's', and best practice advice from organisations active or leading in RDM and related topics. Standards and guidelines are often attached to institutional policies. A national framework might facilitate the standardisation and reuse of locally developed guidelines, and raise awareness of existing best practice advice.

What is the definitive list of standards which a national framework will need to address, including international standards (e.g. GDPR), funder requirements, national regulatory standards (e.g. information standards), state legislation and standards regarding data retention and disposal, other more local policies and standards, and standards required for specific kinds of data (e.g. personal information, sensitive data)? How are standards conflicts addressed, especially when research crosses jurisdictional boundaries? What standards apply to specific disciplines and do such standards differ from the normal risk mitigation and management applied to research in general? Where does responsibility within the institution sit for keeping up with amendments and additions to standards, legislation and regulations? What are researchers' responsibilities and expectations for general awareness of standards, legislation and regulations? Where should expertise and detailed knowledge on these issues sit? Should there be a single institutional entry point for advice and guidance on these issues? How can researchers know when they have done enough to meet a standard and who might check for compliance with standards?

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



[4] "Australian Code for the Responsible Conduct of Research", 2018. https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018