Management of Data and Information in Research

A guide supporting the Australian Code for the Responsible Conduct of Research

2018
Management of Data and Information in Research

A guide supporting the *Australian Code for the Responsible Conduct of Research*
Table of contents

1. Introduction 3
2. Definitions and Significance of Terms 3
3. Responsibilities of institutions 5
   3.1 Control and Ownership of Research Project Information 6
   3.2 Storage, retention and management 7
   3.3 Safety, security and confidentiality 7
   3.4 Access by interested parties 7
   3.5 Facilities 8
4. Responsibilities of researchers 8
   4.1 Selection and publication of Data Ouputs 9
   4.2 Managing confidential information 10
   4.3 Acknowledging the use of others’ data 11
Appendix 12
1. Introduction

This guide complements the *Australian Code for the Responsible Conduct of Research* (the Code), which articulates the broad principles and responsibilities that underpin the responsible conduct of Australian research.

In particular, this guide is intended to assist institutions and researchers to adhere to the following principles of the Code:

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P2 Rigour</strong></td>
<td>in the development, undertaking and reporting of research, which requires that research be characterised by attention to detail and robust methodology and that researchers avoid or acknowledge biases.</td>
</tr>
<tr>
<td><strong>P3 Transparency</strong></td>
<td>in declaring interests and reporting research methodology, data and findings, which requires researchers to share and communicate research methodology, data and findings openly, responsibly and accurately</td>
</tr>
<tr>
<td><strong>P7 Accountability</strong></td>
<td>for the development, undertaking and reporting of research so as to comply with relevant legislation, policies and guidelines and ensure good stewardship of public resources used to conduct research.</td>
</tr>
</tbody>
</table>

The responsible conduct of research includes within its scope the appropriate generation, collection, access, use, analysis, disclosure, storage, retention, disposal, sharing and re-use of data and information. In addition to the principles of the Code, institutional policies that are developed to govern the conduct of research require proper attention to ethics guidelines, privacy legislation and guidelines, other relevant laws, regulations and guidelines, as well as research discipline-specific practices and standards and models for best practice.

2. Definitions and Significance of Terms

This guide deals with three kinds of data and information, which are referred to collectively as *Research Project Information*. *Research Project Information* includes

<table>
<thead>
<tr>
<th>Research Data and Source Materials</th>
<th>Research Data and Source Materials include data and information generated by, collected or accessed for, or used in a research project, including primary research records such as laboratory notebooks, transcripts of interviews or notes of observations. Research Data and Source Materials may also include ‘raw materials’ such as geological samples, soil samples or biological material, or physical or digital objects such as artefacts, questionnaires, sound recordings or video that are the source of data or information used in the project.</th>
</tr>
</thead>
</table>
| Data Outputs                      | Data Outputs are digital data, techniques, algorithms and software that fall into two categories:  
  • those that support the findings of the research and that may have considerable value to other researchers  
  • those that do not necessarily support the findings of the research, |
but may have considerable value to other researchers and the wider community.

Data Outputs may include Research Data and Source Materials, but usually involve additional manipulation, processing, transformation, development and/or intellectual input.

The submission of Data Outputs, especially in raw form, may be a prerequisite for publication or dissemination of the outcomes of research.

Data Outputs are the candidates for discovery, access and sharing.

<table>
<thead>
<tr>
<th>Research Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Records are administrative records that consist of information about a research project (e.g. financial information) that have not been generated by or used in research activity.</td>
</tr>
</tbody>
</table>

Figure 1. The different kinds of information commonly associated with a research project, collectively referred to as Research Project Information.
While it may not be practical to retain all Research Data and Source Materials due to cost, size, scale or format (such as geological samples, biological material, questionnaires or recordings), practical and durable records derived from them (such as assays, test results, transcripts, and laboratory and field notes) must be retained. Such records must be accessible, if required, for verification of the research.

Data Outputs have significant value for verifying and justifying the outcomes of research. They may also have value for the research community and, indirectly, for the public through access and re-use in further research, including in innovative ways that were not apparent at the time of their creation.

In addition, while desirable in principle, it may not be practical to share all Research Data and Source Materials. Factors that are relevant include obligations of confidentiality, intelligibility, and the scale and quality of the data or source material.

Consequently, researchers and institutions should select and prepare the Data Outputs that can be made available and shared openly in accordance with accepted standards for information management such as those endorsed by the National Archives of Australia.¹

### 3. Responsibilities of institutions

Research institutions have a responsibility to develop and implement policies and provide facilities and processes for the safe and secure storage and management of Research Project Information in order to

- allow for the justification and verification of the outcomes of research
- maximise the potential for future research
- minimise waste of resources of value to researchers and the wider community.

This responsibility derives from R1, R2, R3, R7 and, most directly, from R8 of the Code.

| R1 Establish and maintain good governance and management practices for responsible research conduct. |
| R2 Identify and comply with relevant laws, regulations, guidelines and policies related to the conduct of research. |
| R3 Develop and maintain the currency and ready availability of a suite of policies and procedures which ensure that institutional practices are consistent with the principles and responsibilities of the Code. |
| R7 Support the responsible dissemination of research findings. Where necessary, take action to correct the record in a timely manner. |
| R8 Provide access to facilities for the safe and secure storage and management of research data, records and primary materials and, where possible and appropriate, allow access and reference. |

This objective is also consistent with the F.A.I.R. principles, which require that data and information used in research be findable, accessible, interoperable and re-usable.²

---

Research institutions vary in size, maturity, experience and organisational structure. They range from large and complex universities, to small privately funded institutes. Diverse research disciplines also have a range of standards and conventions that need to be considered. Accordingly, this guide acknowledges that different institutional policies and processes are capable of fulfilling the requirements of the Code, and includes options to provide flexibility in its application.

Institutional policy should include guidance for managing Research Project Information that addresses the following:

- Stewardship and ownership
- Storage, retention and disposal
- Safety, security and confidentiality
- Access by interested parties.

The policy/ies should apply to all research conducted under the auspices of the institution.

The policy/ies must be consistent with relevant laws, regulations and guidelines, as well as research discipline-specific practices and standards.

3.1 Control and ownership of Research Project Information

Institutional policy should provide guidance regarding control over Research Project Information, both during the project and after the project has been finalised. This control can be understood as rights equivalent to custody, stewardship or ownership.

Custody, stewardship or ownership of Research Project Information allows the entity asserting those rights, such as an institution, to control access to it, as well as conferring upon it the responsibility to properly manage and store that information.

Ownership of Research Project Information can be difficult to determine, especially when the research involves multiple researchers, externally sourced data or information, and funding or contractual agreements. Institutional policy should clarify the criteria that will be used to determine rights of control of the Research Project Information in these circumstances.

Institutional policy should cover cases where researchers move between institutions or employers and where Research Data and Source Materials or Data Outputs are held outside of Australia. Agreements covering control or ownership of this data or information should be reviewed whenever there is movement or departure of research staff.

An institution may choose to assert ownership of Research Project Information, or to grant ownership to researchers. Institutional policy should emphasise the need to consider funding arrangements for the project that may influence the determination of rights of control or ownership.

As a general rule, the most satisfactory arrangement will be that the materials and data retained at the end of a project are the property of the institution that hosted the project, another institution with an interest in the research, or a central repository.

Institutional policy on these matters should not impede the normal use of Research Data and Source Materials and Data Outputs by researchers for research and scholarly purposes, including their
sharing and communication.

Upon the determination of ownership of Research Data and Source Materials and Data Outputs, a relevant licence should be applied to clarify the status of this data and information with respect to re-use by third parties.

In some instances, research may be conducted partly or entirely using Research Data and Source Materials that are owned by another party. In such cases, neither the institution nor the researcher can assert ownership. Therefore, to meet the requirements outlined in this guide, the researcher and/or the institution will need to document and reference the source of the Research Data and Source Materials and describe the access arrangements. Such arrangements will also need to be in place to allow justification and verification of the outcomes of research.

3.2 Storage, retention and disposal

Institutional policy should address the storage, retention and disposal of all Research Project Information, whether held in an institutional repository or externally.

The storage, retention and disposal of Research Project Information should be consistent with any copyright or licensing arrangements that are in place, as well as complying with relevant privacy and ethical requirements, other relevant laws, regulations and guidelines, and research discipline-specific practices and standards.

This includes clear indications of planned retention periods, which can range from 12 months to 5, 7 or 15 years or more, depending on the nature of the research project. Adherence to minimum requirements for retention of data and information should also reflect a recognition that plans may be required for extremely long-term retention of, for example, observational data from longitudinal studies or sound recordings from endangered languages.

Institutional policy should clarify the requirements for short and long term storage of Research Project Information and how any disposal of data and information is to be undertaken and recorded.

Good archival practice includes scheduled review of items in long-term storage.

3.3 Safety, security and confidentiality

Institutional policy should address the need to maintain the safety, security and confidentiality of Research Project Information during and beyond the lifecycle of a research project. The policy must be consistent with privacy laws\(^3\), and other relevant laws, regulations and guidelines.

3.4 Access by interested parties

Institutional policy should address how to make Data Outputs available to interested parties both within and outside of the institution, giving particular consideration to licensing and access arrangements. Institutional policy should address both open access and mediated access (i.e. access to data or information with the assistance of a data custodian or other authorised person) options for the sharing of Data Outputs.

To the extent possible, institutional policy should promote open access to Data Outputs. In cases where it is not feasible or appropriate for Data Outputs to be shared, alternative processes should be clarified. In most cases, the confidentiality of Data Outputs can be safeguarded by providing a description of the Data Outputs, while regulating access to the Data Outputs themselves (e.g. through mediated access).

Licencing provides a standardised way for researchers and institutions to share Data Outputs with others without infringing copyright. A licence should set out the uses that may lawfully be made of the Data Output and specify the conditions with which use of the Data Output must comply. When considering licensing for use of Data outputs, the least restrictive option, such as a Creative Commons Attribution licence, should be the default.

Institutional policy should also address approved protocols for publication, if relevant, and any requirements related to keeping records of its data assets.

3.5 Facilities

Research Project Information controlled by the institution and/or its researchers should be stored in facilities provided by or approved by the institution. These facilities, including computer systems, must comply with privacy requirements and other relevant laws, regulations and guidelines, and research discipline-specific practices and standards related to safe and secure storage of data and information.

Research institutions offering facilities for the storage of personal or sensitive data or information should have a policy describing their responsibilities as data custodians for the security of and access to data and information.

The policy should also address control or ownership of data facilities and archives in which data or information is stored.

Data Outputs may be published in international, national, or discipline-based repositories, such as international databanks, in addition to institution-based storage or archiving. Institutions should consider maintaining a register of Data Outputs that includes relevant Data Outputs that are held in facilities within and outside of the institution.

4. Responsibilities of researchers

Researchers must adhere to their institution’s policy/ies related to management of data and information, relevant laws, regulations and guidelines, and research discipline-specific practices and standards.

The responsibilities of researchers with respect to management of data and information in research should be clear from the beginning of a research project. The development of a data management plan for this purpose is strongly encouraged and its scope and content are described in the National Statement on Ethical Conduct in Human Research and can be adapted as appropriate.

The guidance provided in this section should be read in conjunction with section 3 (Responsibilities

---


5 https://www.nhmrc.gov.au/guidelines-publications/e72
If not otherwise clarified in institutional policy, researchers are also obliged to

- keep clear and accurate records of research methods and sources of data and information
- exercise adequate care and preservation of primary research records
- ensure access to Research Data and Source Materials, as appropriate, and, in particular, to enable timely response to any challenges. This access should be facilitated by the use of indexes or catalogues of data and information generated, accessed and used during the research
- respect any project-specific conditions of consent or confidentiality requirements
- adhere to project-specific protocols that require measures beyond those required by institutional policy or relevant laws, regulations and guidelines, or research discipline-specific practices and standards
- report any inappropriate use of or access to or loss of data, in accordance with applicable requirements.

In order to optimise project efficiency and avoid information loss and duplication, researchers should employ good management practices. These practices vary across disciplines, but the essential elements include:

- stable storage formats and regular backup to a source external to an individual computer
- version control and other relevant mechanisms for datasets, algorithms, models and software configuration management
- workflow documentation with provenance information for instruments (use and calibration) and software used
- adherence to appropriate national and international standards for scientific terminology and information encoding.

Research Project Information can be the subject of Freedom of Information requests, and in such circumstances, there is an expectation that any information that is delivered will be provided in an understandable format and state.

### 4.1 Selection and publication of Data Outputs

The central aims of retention of data and information are to enable the justification of outcomes of the research and to defend them if they are challenged and the facilitation of sharing of Data Outputs with the research community.

Researchers have primary responsibility for deciding which Research Data and Source Materials are candidates for long-term retention and wider accessibility. Researchers should decide which materials are best able to communicate and support the findings of the project and arrange for appropriate dissemination of these as Data Outputs.

In addition to legal requirements and the requirements of funders, government bodies and publishers, the following criteria should be considered in selecting such outputs:

- Uniqueness and non-replicability
- Reliability, integrity, and usability
- Relevance to a known research initiative or collection
- Social or historical value
- Economic benefit.

Researchers should consider appropriate approaches to maximising the benefits of valuable data and information in the context of any required or reasonable restrictions on sharing the data or information.

In addition to standard publication requirements, options for researchers include publishing or making their Data Outputs available through data centres, national and international collections, or through online repositories maintained by institutions and research communities.

Researchers should adhere to established national and international standards for data description and structuring to facilitate tracking of references. These standards include using Digital Object Identifiers for datasets, ORCID IDs for researchers, and standard terminology for scientific concepts.

Published Data Outputs generally require some kind of online description (i.e. metadata). It is best practice for Data Outputs to be F.A.I.R.—findable, accessible, interoperable, and re-usable both manually and with automated tools. This requires researchers to include appropriate context (descriptive, technical, methodological, access, and provenance information) either within the data structure or in separate metadata records for the Data Output.

Researchers should consider the options for licensing of Data Outputs in order to provide clear parameters around the use and re-use of these outputs.

4.2 Managing confidential information

Researchers must exercise care in handling confidential information used in or arising from a research project. Research Project Information to which obligations of confidentiality may apply commonly fall into one of the following categories:

- Sensitive information (including human medical/health and personal data or information, information about secret or sacred practices, or ecological data that may place vulnerable species at risk)
- Information that is related to national security
- Information that is commercial-in-confidence.

Researchers must ensure that the security and privacy measures\(^6\) that are used for Research Data and Source Materials are proportional to the risks associated with the confidentiality of Research Data and Source Materials. These measures relate to storage, access and sharing of the data and information and should be recorded in the data management plan.

Sensitive Data Outputs may be appropriately shared through mediated access arrangements and the application of a risk assessment framework such as the Five Safes.

\(^6\) Guidance on managing and sharing confidential data and information is provided in the National Statement on Ethical Conduct in Human Research. See Appendix for more information on recommended security and privacy measures.
The Five Safes are:

1. *Safe projects*: is the intended use appropriate?
2. *Safe people*: can the user(s) be trusted to use the information appropriately?
3. *Safe settings*: are there appropriate access controls for collaborating or sharing?
4. *Safe data*: is there a confidentiality risk in any data being shared?
5. *Safe outputs*: are individuals identifiable in the Data Outputs?

4.3 *Acknowledging the use of others’ data*

The Code requires that the work of others is appropriately referenced and cited in the presentation, publication or sharing of research. This principle applies to all data and information used as an input to a research project. In referencing and citing the work of others, researchers should follow accepted norms and standards for scholarly literature and can reasonably expect that their work is acknowledged by others.
Appendix

Resources referred to in this guide, or supporting the principles or responsibilities outlined in the guide include:

**Australian Research Data Commons (ARDC, formerly Australian National Data Service (ANDS))**

**Creative Commons**
https://creativecommons.org.au

**National Archives of Australia**

**National Health and Medical Research Council**

**Office of the Australian Information Commissioner**