

# Research Data Management in Practice

1 May 2013

# Contents

1	Introduction .....	1
2	Why Manage Research Data? .....	1
3	Steps in Research Data Management .....	2
3.1	Research Data Curation .....	2
3.2	Pre Research .....	3
3.3	During Research .....	5
3.4	Post Research .....	6
4	References and Further Reading .....	8

# 1 Introduction

The reuse and management of research data are becoming increasingly important as Australian research funders are increasingly following the global trend of lodging publications and data in public repositories.

Publication and re-usability of research data bring great benefits such as research and researcher effectiveness, enhancing the reputation of researchers and institutions, meeting obligations to funders, and compliance with Open Access agendas. To achieve this, it is critical that research data is properly managed from the pre-research planning stages, through to post research completion.

ANDS has commissioned this “Research Data Management Practice Guide” as a practical starting point that focuses on the 'Why' and 'How' of good data and risk management, with plenty of references for further reading for readers who need more detail. It seems likely that Australia's research funders will follow their international counterparts in requiring a research data management plan as part of the research process; this guide aims to suggest the major roles and responsibilities, including those of the individual researcher, the hosting institution and those of a joint nature.

To assist with effective research data management, ANDS currently publishes research data management resources online at <http://ands.org.au/datamanagement/index.html>. These resources are designed for the benefit of both individual researchers and institutions. They provide a high-level institutional framework and a wide range of specific how-to-guides covering all aspects of research data management.

The Practice Guide is aimed at research administrators in the e-research space, providing them with an overview for the planning and operations of sharing research data, thereby creating better opportunities for data re-use. It is acknowledged that no single person or even business unit is responsible for all aspects of research data management and that a collaborative approach is required. In all cases this will involve the researcher/data creator.

This guide has been developed in accordance with the *Australian Code for the Responsible Conduct of Research*, as at: <http://www.nhmrc.gov.au/guidelines/publications/r39>

## 2 Why Manage Research Data?

Research is becoming more data-intensive. The ability to make research data available for further use delivers a series of benefits including individual and institutional reputation (through data citation), better research- and data validation, improved cost-benefit ratios, and compliance with emerging Open Access agendas. Data sharing will further reduce duplication and will enable more targeted future research. It also supports complex, international research projects.

Not all data can be made publicly available. However, most can be shared if appropriate action is taken to ensure that ethical and other concerns are addressed.

Australian research funders are increasingly following the global trend to require research outputs, including data, to be lodged in public repositories. This is best achieved using standardised approaches to data capture, storage, attribution and metadata.

Recent studies have shown that the benefits of freely available Public Sector Information (PSI), which includes publicly funded research data, outweigh the costs. (<http://ands.org.au/resource/cost-benefit.html>).

Effective research data management also plays a vital role in managing research risk. All research is subject to a range of data-related risks such as data loss or corruption, and privacy or copyright breaches. These risks come with significant, potentially catastrophic impacts. Effective research data management can go a long way towards preventing and managing such risks.

Perhaps the most compelling reason for effective research data management is that it directly benefits both researchers and institutions by minimising the risk of loss or deterioration of research findings (including the data).

### 3 Steps in Research Data Management

Figure 1 below outlines the key steps research institutions need to consider in research data management, respectively before, during and after a research project.

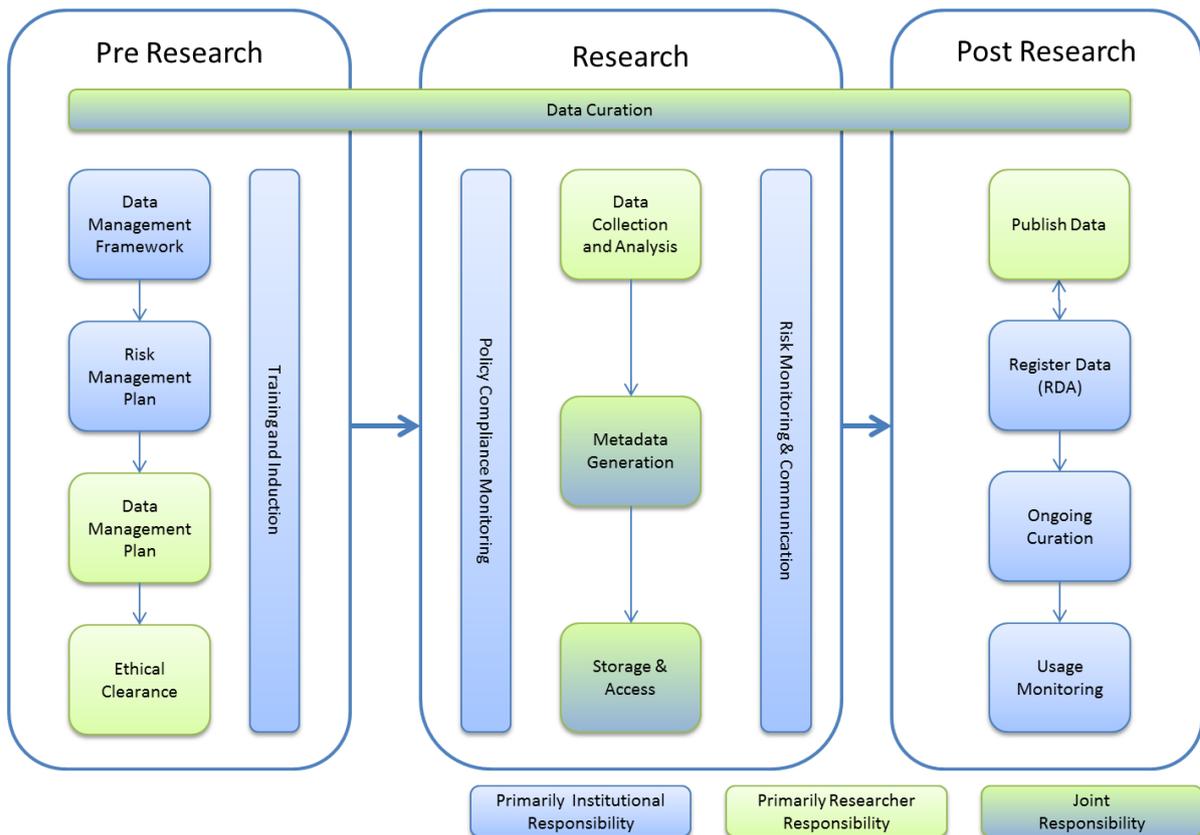


Figure 1 Key steps in Research Data Management

#### 3.1 Research Data Curation

With traditional publication, most curation activities occur at the end of the research cycle. In contrast, digital curation of data is characterised by activities planned for from the outset and occurring throughout the data lifecycle. These will include, for example, the processes outlined in Figure 2.

**Data curation**

- All stages
- Rich metadata
- Access continuum: private, shared, public

Digital data curation requires the capturing of rich metadata (see section 3.3.4), and depositing in appropriate formats into well-managed stores as defined in the Research Data Management Plan (see section 3.2.3)

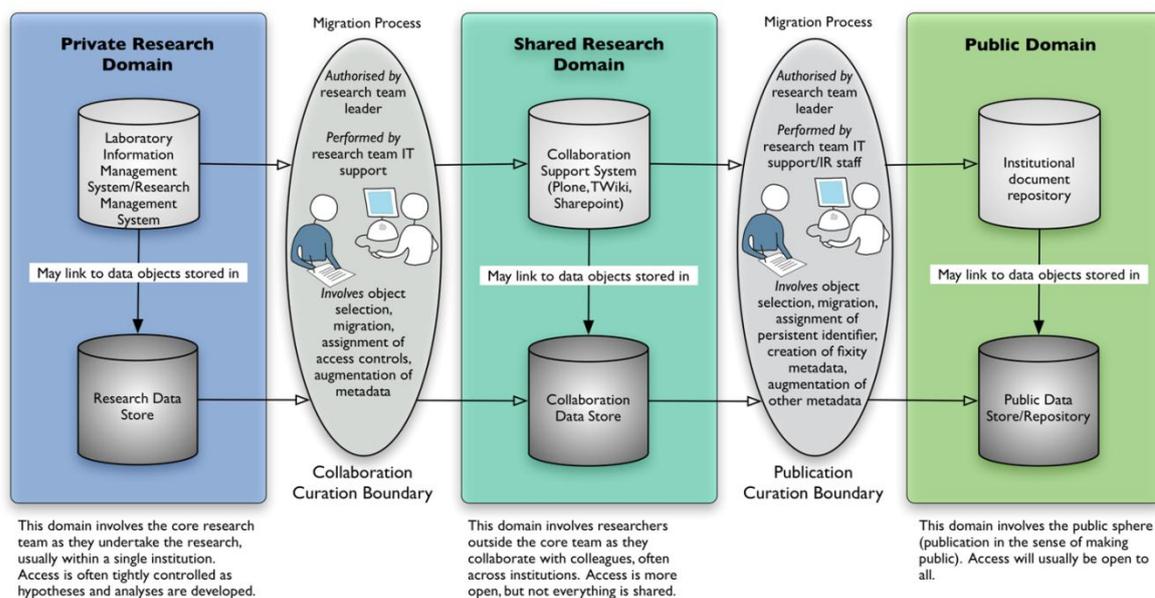


Figure 2 Research Data Curation Continuum. Source: <http://ands.org.au/guides/curation/continuum.html>

## 3.2 Pre Research

### 3.2.1 Research Data Management Framework

The research institution is responsible for providing an adequate Research Data Management Framework, providing the basic elements required within an institutional context to support effective research data management. These elements comprise four categories:

- Institutional policy and procedures: these should be up to date, addressing data-related issues, and be publicised to all those who have a data creation and/or management role;
- IT Infrastructure: the hardware, software and other facilities which underpin data-related activities, as well as identity management and access control;
- Support services: people and other means of providing advice and support, such as web-pages;
- Metadata management: so that data records can be used for both internal and external purposes.



The researcher can benefit from an understanding of the policies, facilities and services of the institution. For more information, see <http://ands.org.au/guides/dmframework/data-management-framework.html>.

### 3.2.2 Risk Management Plan

Risk management is the process of identifying, assessing and responding to risks, and communicating the outcomes of these processes in a timely manner.

Developing a risk management plan is important to avoid preventable losses associated with research data management. Risk can arise from e.g. data loss or corruption, under-utilisation of research outcomes, or breaches of privacy, confidentiality or copyright.



A risk management plan identifies and assesses risks based on likelihood of occurrence and respective impact. It identifies effective mitigation measures based on risk criticality, which are allocated to specific people, and monitored throughout the research project.

An effective risk management plan adheres to the international risk management standard AS/NZS ISO 31000:2009. An example of a guide that builds on this standard is provided by the Queensland Government: <http://www.treasury.qld.gov.au/office/knowledge/docs/risk-management-guide/guide-to-risk-management.pdf>.

An example of an institutional research data risk management plan, including illustrations of risk mitigation strategies, by the Alfred Campus in Melbourne can be found at: [http://www.alfredresearch.org/policies/RMP\\_Research180504.pdf](http://www.alfredresearch.org/policies/RMP_Research180504.pdf)

### 3.2.3 Research Data Management Plan

Development of a Research Data Management Plan is a critical aspect of the pre-research stage. It lays out what data will be created, what policies will apply to the data, who will own and have access to the data, what research data management practices will be used, what facilities and equipment will be required, and who will be responsible for each of these activities.

It will include activities such as:

- data organisation and storage;
- metadata standards and guidelines;
- backups;
- archiving for long-term preservation;
- version control and derived data products;
- data sharing or publishing intentions, including licensing through AusGOAL (<http://www.ausgoal.gov.au>);
- ensuring security of confidential data;
- data synchronisation; and
- governance, roles and responsibilities.



The plan usually defines all research data management related activities during and subsequent to the research activity.

The ANDS Data Management Planning and Awareness Guide provides a detailed checklist of topics to be covered, as well as links to sample plans: <http://www.ands.org.au/resource/data-management-planning.html>

### 3.2.4 Ethical Clearance

Sharing of research data that relates to people can often be achieved using a combination of obtaining consent, anonymising data and regulating data access.

Research data-even sensitive and confidential data-can be shared ethically and legally if researchers pay attention, from the beginning of research, to four important aspects:

- including provision for data sharing when gaining informed consent
- protecting people's identities by anonymising data where needed
- considering controlling access to data



- applying an appropriate licence

More information: <http://ands.org.au/guides/ethics-working-level.html>

### 3.2.5 Training and Induction

Effective implementation of an institutional research data management framework requires that all institutional staff receive adequate training. All staff involved in the research project would benefit from up to date training.

Training and induction

- All staff
- Induction
- Refresher

Furthermore, for project specific issues such as risk management and the implementation of the research data management plan, all researchers and support staff will need to be inducted. Inductions take place not only at project commencement, but also for any personnel coming on board during the project. For longer projects, refresher training and inductions should be considered.

ANDS provides guides, support and training services online, as well as webinars and custom courses. More information can be found at <http://ands.org.au>.

## 3.3 During Research

### 3.3.1 Policy Compliance Monitoring

Acknowledging the policies and guidelines as defined in the institution's research data management framework at the start of a research project is of little use in itself. Demonstrating compliance through review or audit frameworks allows non-compliance to be identified early and corrective action to be taken. This allows the organisation to respond to compliance breaches in a systematic rather than *ad hoc* fashion.

Policy compliance monitoring

- Compliance
- Audits
- Corrective action

Policy compliance is normally dealt with as part of the review of the research data management plan.

### 3.3.2 Risk Monitoring and Communication

Continuous monitoring and review are vital components of an effective risk management process. In terms of the research project, review of risk would normally be incorporated into the review of the research data management plan. The primary purpose of monitoring and review is to determine whether risks still exist, whether new risks have emerged, and to reassess the risk priorities.

Risk monitoring and communication

- Monitor
- Review
- Re-assess
- Communicate

### 3.3.3 Research Data Collection and Analysis

Throughout the research cycle data will be collected. Data analysis will generate derived data and in many cases data will be shared between researchers and institutions for collaboration purposes.

Data collection and analysis

- Sharing with collaborators
- Derivation
- Curation

Data at this point may be made available to collaborators, as specified in the research data management plan.

### 3.3.4 Metadata Generation

Data discovery and access is dependent on the availability of rich metadata. Metadata is collected at both collection and object level and can be stored separately, or embedded in the data collection. Collection level metadata is generated by the researcher as part of the research process and supplemented by object level metadata for data publication, in most cases, by librarians or other data

Metadata generation

- Rich metadata from the outset
- At collection and object level

professionals.

Good metadata creation can be supported by tools designed to simplify metadata input and to enhance interoperability.

Metadata is also useful in tracking the history of derived data products.

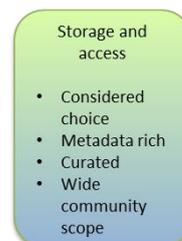
Further reading: <http://www.ands.org.au/metadata/index.html>

### 3.3.5 Storage and Access

The choices made regarding the approach to data storage have implications for cost, security, and future access. It is an institutional responsibility to ensure that adequate and appropriate storage facilities are available.

The goals of "reusing and sharing data more often" are met by storage solutions which make data discoverable and accessible over the long term, which means the tendency should be towards more metadata rich, curated stores with a wide community scope.

More information in the ANDS Data Storage Guide: <http://ands.org.au/guides/storage.html>



Storage and access

- Considered choice
- Metadata rich
- Curated
- Wide community scope

## 3.4 Post Research

### 3.4.1 Publish Research Data

There is an increasing expectation that the outputs of publicly funded research, including the data, will be made available for others to use. That means published data should be well-described (metadata), citable, discoverable and re-usable wherever possible.

Potential reusers of research data need to have clear guidance about what they can and cannot do with the data: this is normally achieved via a licence. ANDS supports the Australian Governments Open Access and Licensing Framework (AusGOAL - <http://www.ausgoal.gov.au/>) which provides support and guidance to facilitate open access to publicly funded information.

More information on copyright and licensing can be found at <http://ands.org.au/publishing/licensing.html>.

Research data can be published in the form of collection descriptions, citable and online accessible data elements, or citable other objects such as web services, APIs, or concept definitions. Sometimes just the existence of data collections is published; this occurs when data cannot be accessed or accessed under strict conditions. Online data publication for download or web-service access is desirable for those kinds of data which are not restricted.

ANDS provides an overview of the many aspects of data publication here: <http://www.ands.org.au/publishing/>.



Publish data

- Well-described
- Citable
- Discoverable
- Re-usable
- Ethics & copyright

### 3.4.2 Register Research Data

The process of registration and citation depends on the type of data published and how it is published. Data can be considered as 'published' when it is generally discoverable. ANDS recommends that registration with Research Data Australia is one form of data publication.

**Research Data Australia (RDA)** is designed to expose the description and existence of data collections. It holds collection description information - information about collections and the activities, parties and services (for example, RSS feeds) which relate to collections.



Register data

- Research Data Australia
- Cite my data
- DOIs
- Identifiers

Note that RDA only registers collection descriptions; it does not store the data.

The **ANDS Cite My Data Service** (<http://www.ands.org.au/services/cite-my-data.html>) allows research organisations to assign Digital Object Identifiers (DOIs) to research datasets or collections. The DOI system supports the citation of research data in scholarly communications and research data collaborations.

Importantly, DOIs also provide the opportunity to measure data reuse and research effectiveness.

Research datasets and collections with a DOI are publicly discoverable.



Figure 3 Using Digital Object Identifiers (source: ANDS)

The datasets or collections assigned DOIs by the Cite My Data service should be citable contributions to the scholarly record, analogous to a journal article. There is a many-to-many relationship between research data and the academic publications it informs.

Other kinds of digital objects (such as web services, APIs, concepts or people) may be better identified using a persistent identifier available via the **ANDS Identify My Data service** (<http://www.ands.org.au/services/identify-my-data.html>).

### 3.4.3 Ongoing Curation

Research data that is not shared and not (immediately) needed for further research should be properly archived, or disposed of at an appropriate time.

The ANDS Data Management Framework (<http://www.ands.org.au/guides/dmframework/dmf-policies-and-procedures.html#Retention>) states: *Universities are guided by the requirements of the Australian Code for the Responsible Conduct of Research which sets out retention periods for research data, research records and primary materials.*

#### Ongoing curation

- Non-shared or used data
- Retention periods
- Regulatory requirements
- Proper disposal

In addition, there is need to consider

- specific provisions set out in Commonwealth and State legislation concerning archives and record keeping, especially those materials of heritage value,
- disciplinary requirements, and
- possibly other regulatory requirements.

### 3.4.4 Usage Monitoring

One of the great benefits of publishing research data with proper identifiers is the ability to track usage and citation statistics. It is recommended that institutions monitor data usage on a regular basis.

#### Usage monitoring

- Citation tracking
- Online tools

Data citation tracking supports research evaluation. The *Data Citation Index* provides data citation metrics. ([http://wokinfo.com/products\\_tools/multidisciplinary/dci/](http://wokinfo.com/products_tools/multidisciplinary/dci/))

## 4 References and Further Reading

Title	URL
ANDS Data Management Framework	<a href="http://ands.org.au/guides/dmframework/data-management-framework.html">http://ands.org.au/guides/dmframework/data-management-framework.html</a>
ANDS Data Management online resources	<a href="http://ands.org.au/datamanagement/index.html">http://ands.org.au/datamanagement/index.html</a>
ANDS Guides and Other Resources	<a href="http://ands.org.au/guides/index.html">http://ands.org.au/guides/index.html</a>
ANDS Guides, Training and Support	<a href="http://ands.org.au/">http://ands.org.au/</a>
AS/NZS ISO 31000:2009. Risk management - Principles and guidelines	<a href="http://sherg.org/31000.pdf">http://sherg.org/31000.pdf</a>
AusGOAL - Australian Governments Open Access and Licensing Framework	<a href="http://www.ausgoal.gov.au/">http://www.ausgoal.gov.au/</a>
Australian Research Data Commons	<a href="http://ands.org.au/ardc.html">http://ands.org.au/ardc.html</a>
Costs and Benefits of Research Data Provision	<a href="http://ands.org.au/resource/cost-benefit.html">http://ands.org.au/resource/cost-benefit.html</a>
Institutional Research Data Management Policies and Procedures	<a href="http://ands.org.au/guides/dmframework/dmf-policies-and-procedures.html">http://ands.org.au/guides/dmframework/dmf-policies-and-procedures.html</a>
Institutional Research Data Risk Management Plan (by the Research & Ethics Unit of the Alfred Medical Research and Education Precinct: <a href="http://www.alfredresearch.org/">http://www.alfredresearch.org/</a> )	<a href="http://www.alfredresearch.org/policies/RMP_Research180504.pdf">http://www.alfredresearch.org/policies/RMP_Research180504.pdf</a>
Queensland Government Risk Management Guide	<a href="http://www.treasury.qld.gov.au/office/knowledge/docs/risk-management-guide/guide-to-risk-management.pdf">http://www.treasury.qld.gov.au/office/knowledge/docs/risk-management-guide/guide-to-risk-management.pdf</a>
Research Data Australia	<a href="http://researchdata.ands.org.au/">http://researchdata.ands.org.au/</a>
Research Data Curation Continuum	<a href="http://ands.org.au/guides/curation.continuum.html">http://ands.org.au/guides/curation.continuum.html</a>