

Data management planning

Who should read this?

This is likely to be of particular interest to researchers and others in the research team who are charged with managing data for a research project or an institution.

What is data management planning?

Data management includes all activities associated with data other than the direct use of the data. It may include:

- data organisation;
- backups;
- archiving data for long-term preservation;
- data sharing or publishing;
- ensuring security of confidential data; and
- data synchronisation.

A *data management plan* is a document that describes what data will be created, what policies will apply to the data, who will own and have access to the data, what data management practices will be used, what facilities and equipment will be required, and who will be responsible for each of these activities.

All Australian universities and research institutions have policies and guidelines relating to different aspects of data management. This may include guidelines on how to construct a data management plan. You should make sure you are aware of such policies and guidelines before you start your plan.

Why do I need a data management plan?

The carrot: Taking time at the start of a research project to put in place robust, easy-to-use data management procedures will usually pay off several times over in the later stages of the project in improvements to efficiency, protection, quality & exposure.

The stick: Basic data management is required by the *Australian Code for the Responsible Conduct of Research*. The *Code* was developed by the ARC, NH&MRC and Universities Australia. A data management plan is now required by some funders at the time of applying for funding

Inadequate preparation for good data management can also lead to catastrophes like the loss of data or the violation of people's privacy.

What does a data management plan need to cover?

The following list of topics can be treated as a check-list:

1. Survey of existing data: What existing data will need to be managed?
2. Data to be created: What data will your project create?
3. Data owners & stakeholders: Who will own the data created, and who would be interested in it?
4. File formats: What file formats will you use for your data?
5. Metadata: What metadata will you keep? What format or standard will you follow?



6. Access & security: Who will have access to your data? If the data is sensitive, how will you protect it from unauthorised access?
7. Data organisation: How will you name your data files? How will you organise your data into folders? How will you manage transfers and synchronisation of data between different machines? How will you manage collaborative writing with your colleagues? How will you keep track of the different versions of your data files and documents?
8. Storage: Where will your data be stored? Who will pay for the hardware? Who will manage it?
9. Backups: This is probably the single most important item on this list. Hard drives on desktop and laptop computers fail regularly. You must have a credible backup strategy of regular backups, and of course you must then follow it. Consider including an off-site backup so that your data will not be lost if your building burns down. Rather than relying on memory, consider an automated backup process.
10. Bibliography management: What bibliography management tools will you use? How will you share references with the other members of your group?
11. Data sharing, publishing and archiving: What data will you share with others? How will you do this?
12. Disposal: What data will you dispose of? When? How?
13. Responsibilities: Who will be responsible for each of the items in this plan?
14. Budget: What will this plan cost? Possible costs include hardware for backups, research assistant time for data curation, metadata creation, archiving etc.
15. Anything else: Don't restrict yourself to the items above. Stop and think. What is missing from this list? (And if you think of something, please let us know so that we can include it in the next version of this document.)

Other issues to consider

Funding bodies and governments are moving rapidly to require sound data management. You have a responsibility to make yourself aware of any relevant codes and to comply with them. Failure to comply with requirements from funding bodies like the ARC or NHMRC may jeopardise future research funding. Failure to comply with legal requirements, such as those that safeguard the privacy of participants in medical research, may lead to prosecution.

Different disciplines have different conventions. In order to facilitate cooperation, you should make sure that your data management is compatible with the prevailing standards in your discipline. (This mostly applies to file formats and metadata standards.)

Further information

ANDS Guides and other Resources: <http://www.ands.org.au/guides/>

ANU Data Management Manual,

http://information.anu.edu.au/training_and_skills_development/information_literacy/resources/ANU_DM_Manual-v11.09.20_v2.pdf

Monash University Research Data Management web site. <http://www.researchdata.monash.edu.au/>

ARC Funding Agreement for Discovery Projects, http://www.arc.gov.au/ncgp/dp/dp_fundingagreement.htm. (In the agreement for 2010 projects, data management requirements are in Section 20, on page 18.)

Australian Code for the Responsible Conduct of Research,

<http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>

Thanks to Mark Euston, who wrote the *ANU Data Management Manual* and *Template* cited above, on which this document is based.

