• Give an overview of research data management and cover the important aspects for researchers to consider

• Provide information on best practices, problems to avoid, and tools available
• Overview
• What is Research Data?
• Why is Research Data Management important?
• Data Management at Murdoch
• FAIR Data
• Data Management Planning
• Tips and Tricks
• Resources and Tools
• Where to get help
What is Research Data?

Research data can be anything that may be needed to validate the results of research.

Research data is a valid form of research output and is increasingly being recognised as a valuable asset which should be managed and, when appropriate, shared.

Some journals now require that data be made available to support research conclusions and the sharing of data or the provision of a data management plan may be required as a condition of research funding.
Research data may include:

- statistical data and analyses,
- measurements,
- questionnaires,
- interview transcripts,
- fieldwork notes,
- images,
- sound/video recordings, and
- artefacts.

The data could be physical or digital; it may be original, transcribed or anonymised.
Good data management allows you to:

- Work more efficiently
- Achieve greater exposure for your research through publication, citation and collaboration
- Improve grant funding opportunities
- Protect your data from misuse or loss
- Enable future researchers to build upon your work
- Comply with the University's Responsible Conduct of Research Policy
Murdoch currently includes policy on Research Data Management in the Responsible Conduct of Research Policy:


There is further detail in the Research Data and Materials Recordkeeping Guideline:


A new Research Data Management Policy is coming soon, and this will involve development of new planning tools and procedures.
The FAIR principles were developed to address one of the challenges of data-intensive science by making data Findable, Accessible, Interoperable and Reusable.

The principles emphasise machine-actionability because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.
FAIR Data

Findable
The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.

Accessible
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

Interoperable
The data usually need to be integrated with other data. In addition, the data needs to interoperate with applications or workflows for analysis, storage, and processing.

Reusable
Metadata and data should be well-described so that they can be replicated and/or combined in different settings.
Data Management Planning

Have a plan before you start – the plan can evolve with the research, but it’s hard to retrofit good planning.

There are questions to consider in the following areas:

- Data Collection
- Documentation and Metadata
- Ethics and Legal Compliance
- Storage and Backup
- Selection and Preservation
- Data Sharing
- Responsibilities and Resources
Data Collection

• What data will you collect or create?

• How will the data be collected or created? What are the best practices?

• What programs and file formats will you use?

• Is there existing data that you could use or might want to use?
Documentation and Metadata

• How will you describe your data?

• What metadata will you keep?

• What format or standard will you follow?

• Will other people be able to understand the data later?
Ethics and Legal Compliance

- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?
- Who will own the data?
- How will you manage collaboration?
- Are there retention rules?
Storage and Backup

• Where will your data be stored? Who will pay for and manage the hardware?

• How will you name, sort and manage the data files? How will you manage transfers and synchronization?

• How will you back up your data?

• How will you manage access and security?
Selection and Preservation

• Which data are of long-term value and should be retained, shared, and/or preserved?

• What is the long-term preservation plan for the dataset?

• What data will you destroy? When? How?
Data Sharing

• How will you share the data?

• Are any restrictions on data sharing required by the type of data, the funder, your co-authors or your institution?

• When sharing the data, what licence will you make it available under?
Responsibilities and Resources

• Who will be responsible for data management?

• What resources will you require to deliver your plan?

• What will the plan cost? Possible costs include hardware, software, data curation time, metadata creation, archiving, etc.
When creating data, describe it as you go – make sure you document sources, methods and reagents so that others will understand.

Store the data in more than one place (cloud vs home), and on more than one medium (Flash drive, hard drive, cloud drive).

Store the data in a format that is machine-readable:
- Don’t merge cells in spreadsheets
- Don’t use cell colours to denote meaning – computers can’t read this. Add a column instead
- Use consistent layouts, labels and terminology
- Store dates in separate columns for day, month and year to avoid format issues

Avoid proprietary formats if you can
- If the software changes, the data can be changed, or made unreadable
- Others may not be able to access the software
Organise your files so that they make sense!

Create a version scheme that makes sense – no finalfinal.xlsx
- Use a number system – v1, v2, v3
- Incorporate descriptions to indicate why the version is different
- Include a date in ISO 8601 format YYYYMMDD
- Status can be included, but be careful

Create good file names
- Be descriptive but concise (filename length can be an issue)
- Create unique file names
- Avoid special characters
- Use underscores instead of spaces (to account for OS changes)
- Use leading numbers for dates so you can sort chronologically
- Remember that some OS are case sensitive
Tips and Tricks

When naming files, try to include:
• Date and time of creation
• Project name or number
• Location
• Name of creator
• Short description
• Version
• Analysis undertaken
• Anything else that will be important for recognizing the file

Seems like a lot, but can be achieved with initials/codes. If this gets complex, use a readme.txt file to record the system!

• E.g. 20190827_RDM_CF_PPT_v2NewFormat.ppt
Always document terminology or acronyms, not everyone uses the same ones

Be careful if using folders to organize files – if you move the file out of its folder, it loses its context

Never have a Miscellaneous folder

Be careful with programs that have auto-correct, especially Excel. This has created serious problems with transformation of genetic data into date formats, for example.
Office365 and OneDrive – 1TB of storage, secured by Murdoch login

Cloudstor – 1 TB of storage, secured by Murdoch login, eResearch tools like Jupyter Notebooks

Always remember that cloud storage is tied to your institutional account – tidy up and back up before you leave!
Resources

Australian Research Data Commons (ARDC) and Australian National Data Service
https://www.ands.org.au/guides

Research Data Australia
https://researchdata.ands.org.au/

Research Data Alliance
https://www.rd-alliance.org/
Resources

Research Data Management Guide

https://libguides.murdoch.edu.au/RDM
Library Help Desk
10am – 2pm Monday to Friday
North Wing Level 3
(opposite library main entrance)
Consultations

- Provide further research support and skills training for academic staff and research students

- You can arrange an individual or small group consultation
Conclusion

Any questions?

Please help us by completing the feedback survey.