MONASH University Library

GROUP OF EIGHT AUSTRALIA

Effective Research Data Management – why it matters

Neil Dickson | Research data management coordinator

researchdata@monash.edu

Learning outcomes

- ☐ Recognise the characteristics of research data and how research data in practice will benefit your research career
- ☐ Identify best practice in managing research data
- ☐ Develop a research data management plan
- ☐ Locate available resources and tools to help you manage your data

Policy and “the Code”

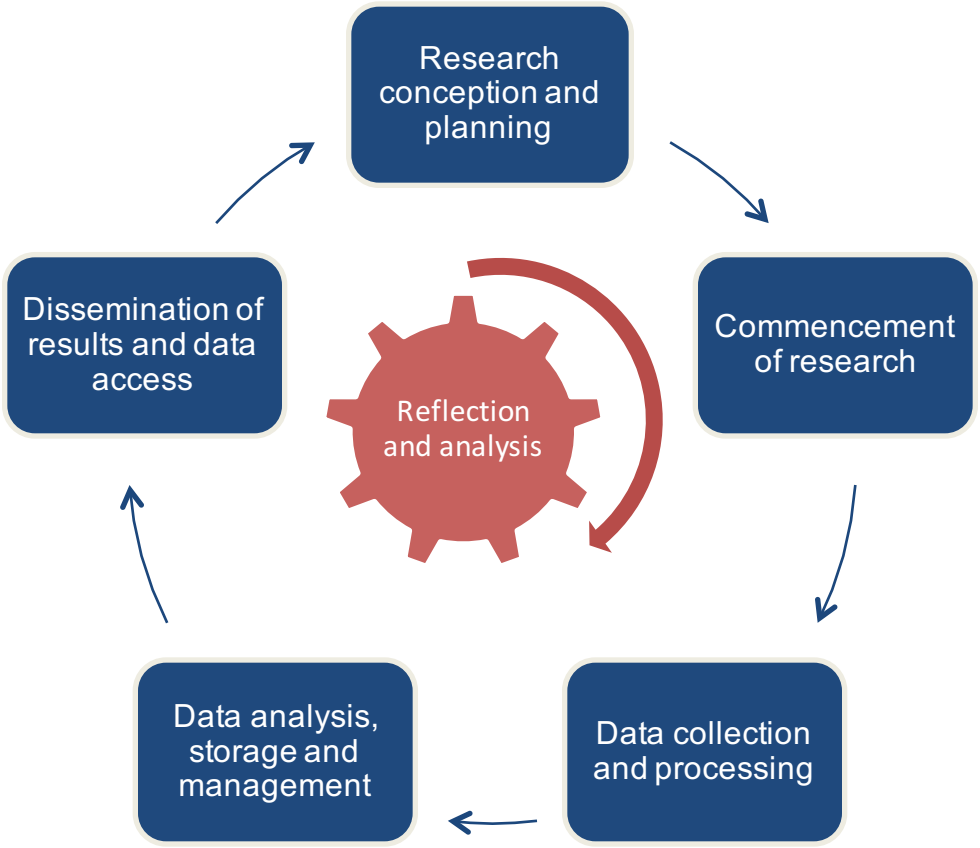
Monash University Research Data Management Policy and Procedures (including HDR procedures)
www.researchdata.monash.edu.au/policies.html

Australian Code for the Responsible Conduct of Research (2007), Section 2
www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/r39.pdf

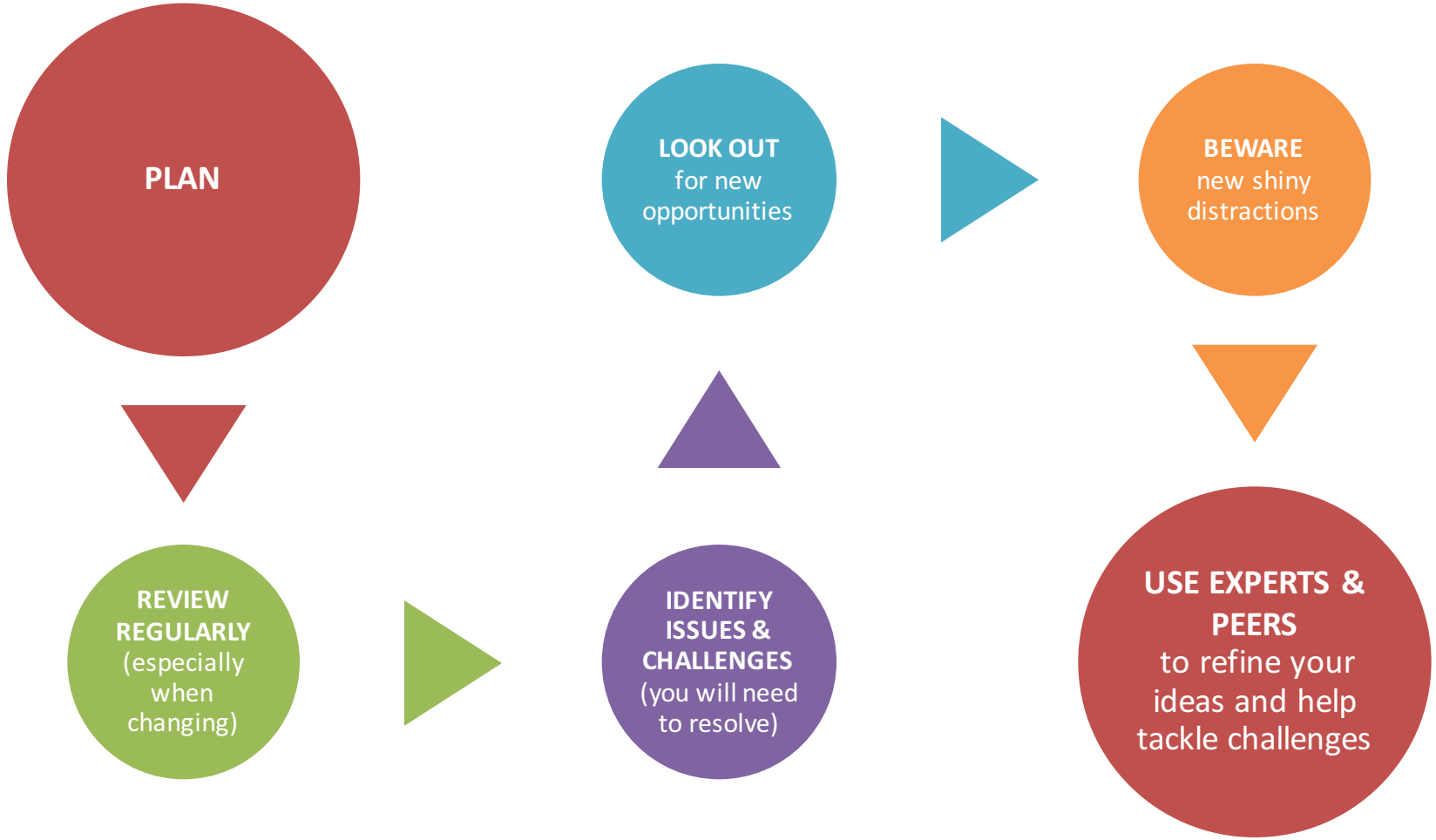
“ The responsible conduct of research includes the proper management and retention of research data.

The central aim is that sufficient materials and data are retained to justify the outcomes of the research and to defend them if they are challenged. The potential value of the material for further research should also be considered. ”

Research data cycle



Key points



MONASH University Library

Challenges of managing research data

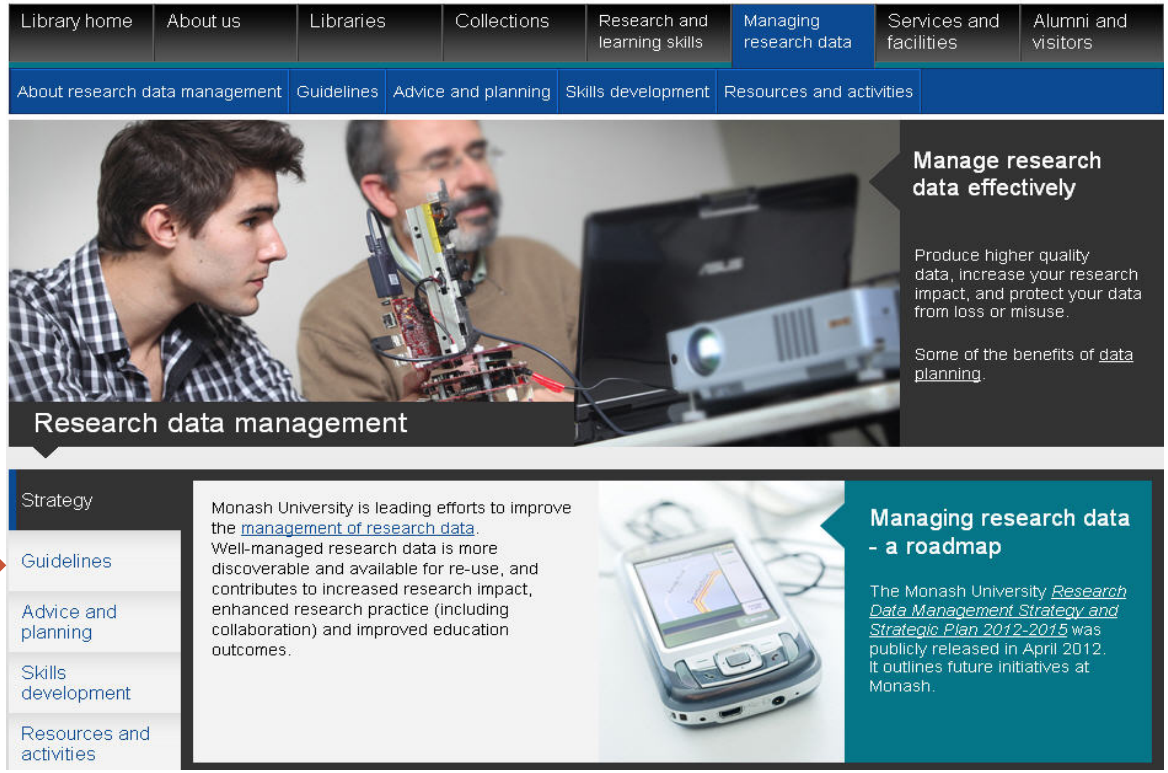
- Access to knowledge and information
- Planning
- Ownership copyright and IP
- Ethics, privacy, confidentiality and cultural sensitivity
- Ethics and data sharing
- Anonymising qualitative data
- Anonymising quantitative data

- Security
- Storage needs
- File types
  - Obsolescence
  - Interoperability
  - Discipline standards/norms
- Sharing and collaboration
- Data retention
  - Describe your data (metadata)
  - Organise your data (folder & file naming)

Managing research data – Library website

http://monash.edu/library/researchdata/index.html

Start here



MONASH University Library

Why create a data management plan?

- Consider all aspects of research data management early in your project
- Identify areas of potential difficulty or conflict
- Resolve these with your supervisor and other Monash staff (e.g. Library, e-Research Centre *MeRC*)
- Find out about available services and tools
- Some kind of plan may be associated with confirmation of candidature in future
- Some kind of plan may be a requirement of future grant submissions

What to do with your planning checklist

http://www.researchdata.monash.edu.au/resources/datahdrchecklist.doc

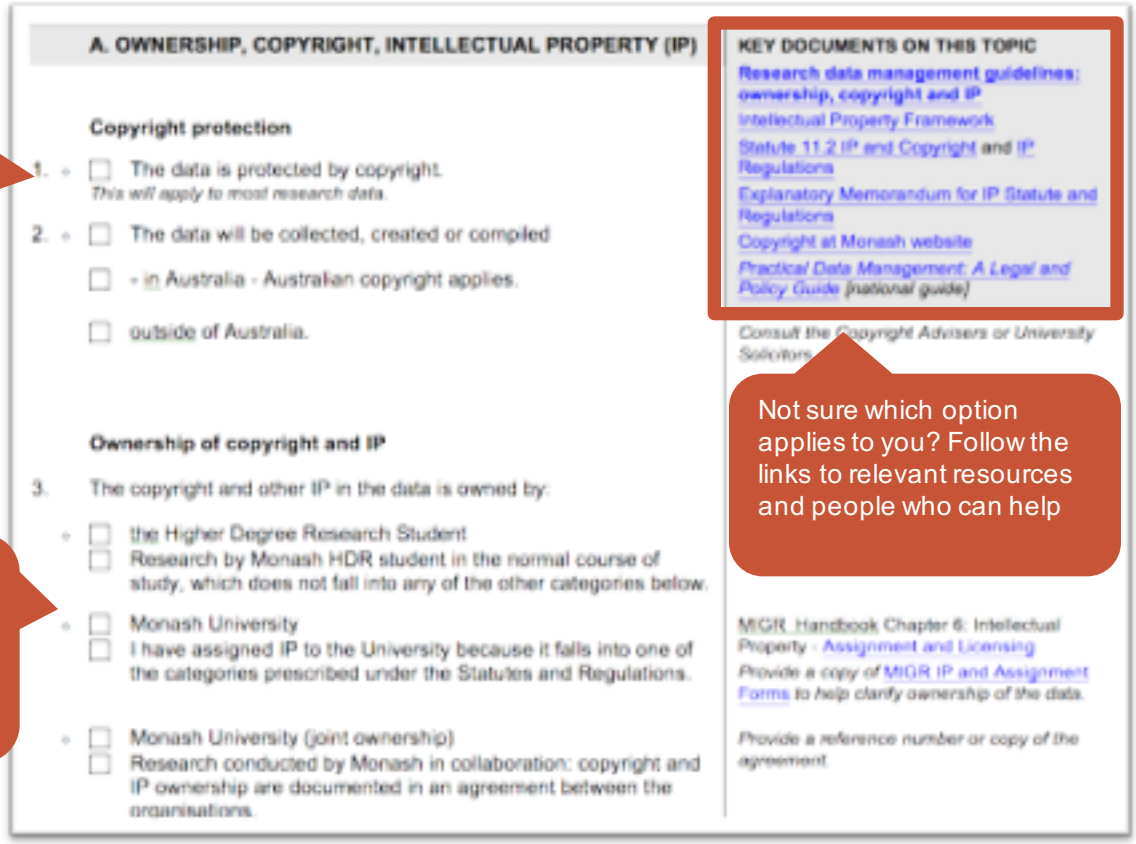
- Keep a copy for yourself
- Use it as a conversation starter with your supervisor and other staff and peers
- Review it regularly as your project progresses – your approach to managing your data will evolve along with your research

HDR Data planning checklist

www.researchdata.monash.edu.au/resources/datahdrchecklist.doc

Complete quickly and easily using multi-choice boxes

Attach other documents and add supplementary information to create a more comprehensive data management plan



A. OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP)

Copyright protection

1. ☐ The data is protected by copyright.
   This will apply to most research data.
2. ☐ The data will be collected, created or compiled
   ☐ - in Australia - Australian copyright applies.
   ☐ outside of Australia.

Ownership of copyright and IP

3. The copyright and other IP in the data is owned by:
   ☐ the Higher Degree Research Student
   ☐ Research by Monash HDR student in the normal course of study, which does not fall into any of the other categories below.
   ☐ Monash University
   ☐ I have assigned IP to the University because it falls into one of the categories prescribed under the Statutes and Regulations.
   ☐ Monash University (joint ownership)
   ☐ Research conducted by Monash in collaboration: copyright and IP ownership are documented in an agreement between the organisations.

KEY DOCUMENTS ON THIS TOPIC

Research data management guidelines: ownership, copyright and IP
Intellectual Property Framework
Statute 11.2 IP and Copyright and IP Regulations
Explanatory Memorandum for IP Statute and Regulations
Copyright at Monash website
Practical Data Management: A Legal and Policy Guide (national guide)

Consult the Copyright Advisers or University Solicitors

MGR Handbook Chapter 6: Intellectual Property - Assignment and Licensing
Provide a copy of MGR IP and Assignment Forms to help clarify ownership of the data.

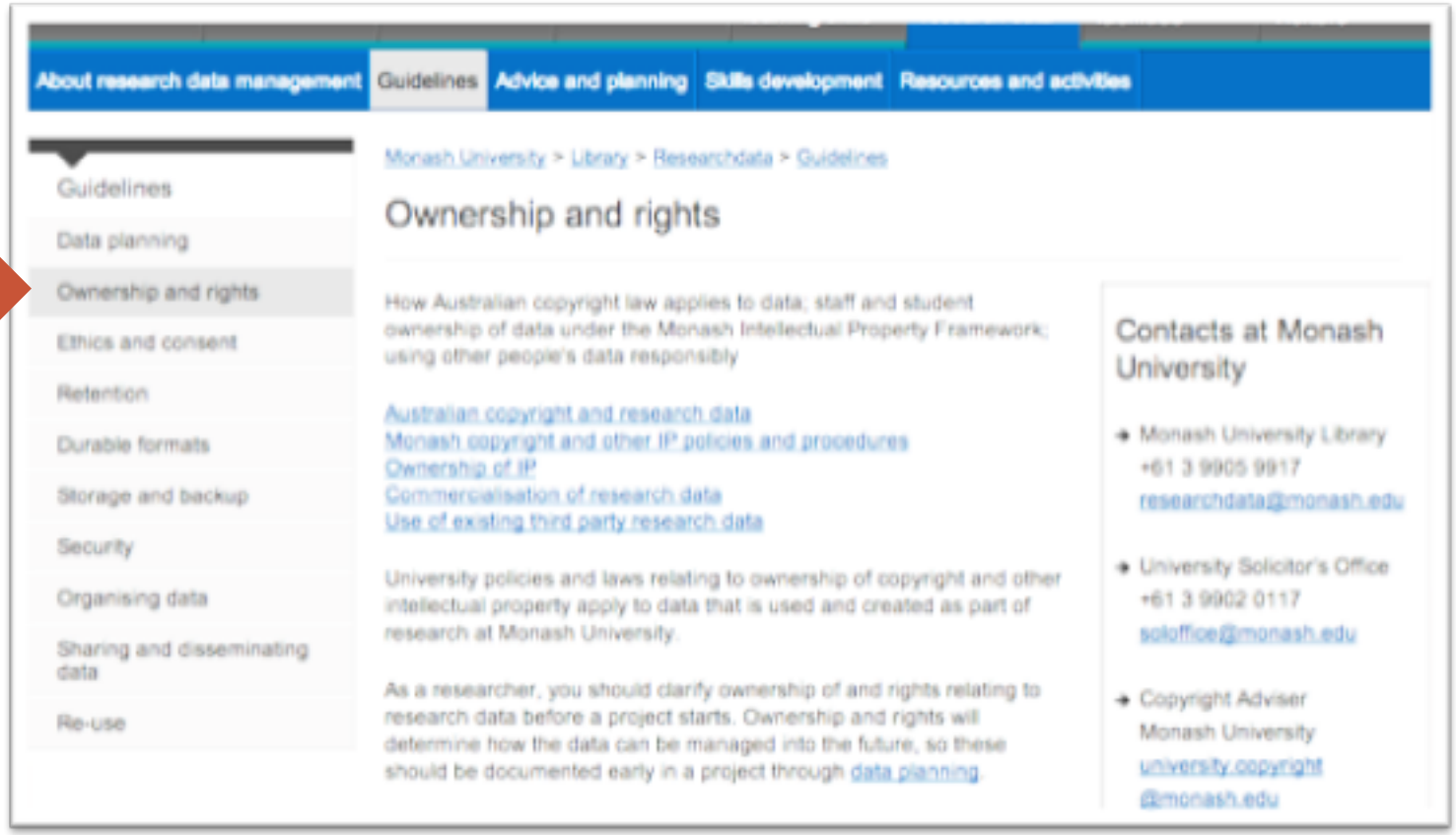
Provide a reference number or copy of the agreement

Not sure which option applies to you? Follow the links to relevant resources and people who can help

MONASH University Library

Ownership Copyright and IP

http://monash.edu.au/library/researchdata/guidelines/ownership/



MONASH University Library

Copyright basics

- A type of intellectual property, like trademarks or patents
- Owners have exclusive rights to do things or authorise others to do things with protected material
- Protects the material expression of an idea, not the idea itself
- Does not need to be registered in Australia – it applies automatically
- If your research is conducted in Australia, then Australian copyright law will apply
- In Australia copyright will apply to data in almost all cases - 'originality' only requires that some skill, ingenuity and labour has been involved, not that the thought or idea is particularly novel

Data you create or collect

The Monash Intellectual Property Framework applies to data as well as to your thesis

- In general, students own copyright and IP in any data they generate, but there are exceptions
- The IP Assignment and Deed of Assignment of Intellectual Property forms clarify these issues

Using third party data

Any data you use that is not of your own creation may be protected by copyright

You don't need to do anything if

- The data is out of copyright, or
- Your usage falls within the scope of the 'fair dealing' provisions in the Copyright Act

In all other situations you need to have permission to use the data

'Express permission' - usually found in Terms and Conditions of a purchase agreement or licence

If there is no express permission, you need to approach the copyright owners yourself

Just because data is available for free on the web does not mean there are no terms and conditions associated with its use

It is your responsibility to be clear about what you can and can't do

Supporting documents for third party use

- Print-out of owner's statement of express permission
- Copies of the written permissions (e.g. emails, letters) owners have provided in response to your requests

Ownership, copyright and IP: where to go for detailed information

- *Copyright at Monash* website
  http://www.copyright.monash.edu.au/
  - guidelines on all aspects of copyright, including third party content and special types of materials
  - templates for seeking permissions from owners

- University Copyright Advisers
  University.Copyright@lib.monash.edu.au

- *Practical Data Management: A Legal and Policy Guide* (2008)
  http://www.oaklaw.qut.edu.au/reports

Durable formats: considerations

Will it last at least for the lifetime of the project and the minimum retention period, or be able to be converted to a new format?

For digital data, is the format

- endorsed and published by a standards agency e.g. Standards Australia or ISO?
- independent from specific hardware or software?
- widely used and accepted as best practice within your discipline or user community?

Required hardware/software

- How long will the hardware/software be available?
- How widely used is it?
- What level of support is available now, and likely to be available in the future?
- If licensing allows, you might want to keep a copy of the software and documentation (e.g. user manual) with your digital data

3-2-1 strategy for backups

3 backups

2 backups stored by you in your own storage

1 backup stored in Monash managed storage

Digital data considerations and network drives

Monash network storage/drives

- Storing your research data on network drives is **highly recommended** benefits include:
- all your research data in a single place
- automatic backups and integrity checks
- data is readily available to authorised users, including via remote access if needed
- standard security and access controls

Other storage solutions

- How likely is this technology to fail or becoming obsolete?
  - What will be the impact if it does?
- How long do media usually last?
  - Under what environmental conditions?
- What security, backup and disaster recovery procedures are in place?
- Is professional IT support available?

Network drives and storage options

http://monash.edu/library/researchdata/file_links/storage_options_web_vers15_10_2013.pdf

Research data storage options at Monash University

Decision points							Storage options						
Are you working with [1]very large datasets?	Are you using [2]active data?	Are you using the Monash Windows [3]SOE?	Are you using [4]sensitive data?	Are you sharing your data?	Are your research partners external to Monash?	Do you require remote access?	**S: Drive**	**Local drives** (desktop, My Documents etc.)	**Monash's Google apps** (Docs, Spread sheets, files in Google Drive)	**LaRDS** (and other managed research data storage)	**Customiseds solution** (case by case individual basis)	**Transportable storage devices** (flash drives, thumb drives storage etc.)	**AARNet's CloudStor service** (data transfer only)
[1]Yes	ⓘ Seek advice at this point ➔									✓ eSolutions Service Desk(SDO) For referral			
No	No	ⓘ Seek advice at this point ➔								✓ Contact the Library For referral			
No	Yes	No	ⓘ Seek advice at this point ➔ For non-Windows SOE users (including Mac users) available storage options include those ticked. A suitable encryption tool is TrueCrypt (refer to tutorial). For non-standard or complex data storage arrangements, seek advice from the eSolutions Service Desk (SDO).					✓	✓	✓ eSolutions Service Desk(SDO) For referral		✓	✓ CloudStor service
No	Yes	Yes	Yes	Yes	Yes	Yes					✓ eSolutions Service Desk(SDO) For referral		
No	Yes	Yes	Yes	Yes	No	Yes	✓ eSolutions Service Desk(SDO) Multiple owners recommended					✓ Encrypt device (BitLocker to Go)	
No	Yes	Yes	Yes	Yes	No	No	✓ eSolutions Service Desk(SDO) Multiple owners recommended					✓ Encrypt device (BitLocker to Go	
No	Yes	Yes	Yes	No	No	No	✓ eSolutions Service Desk(SDO) Multiple owners recommended						
No	Yes	Yes	No	Yes	Answer can be **Yes or No**	Answer can be **Yes or No**	✓ eSolutions Service Desk(SDO) Multiple owners recommended		✓			✓	✓ CloudStor service
No	Yes	Yes	No	No	No	Yes	✓ eSolutions Service Desk(SDO)		✓			✓	✓ CloudStor service
No	Yes	Yes	No	No	No	No	✓ eSolutions Service Desk(SDO)	✓	✓				

1. Dataset size (guide only): 'small', up to 30GB; 'medium', between 30GB and 100GB; 'large', greater than 100GB up to 500GB; and, 'very large', greater than 500GB.
2. For the purposes of this document, active or working data are defined as data that require ongoing access for modification, analysis, compilation, etc. Archival storage solutions are more appropriate for 'end state' data.
3. Monash recommends use of the Monash Windows Standard Operating Environment (SOE) and can only provide limited support for data stored on non-SOE and non-Monash controlled environments. Mac users - the Mac OSX SOE is not yet certified.
4. For "sensitive" data classifications, refer to the following legislative definitions: Information Privacy Act 2000 (Vic), Privacy Act 1988; and, the Monash interpretation in the Electronic Information Security: Responsibilities, Classifications and Standards Procedures.

MONASH University Library

Monash storage options – benefits vs limitations

Research data storage options at Monash University

	Benefits	Limitations
S: Drive • For Monash staff and students: requires authcate access • Networked share drive: replacement for faculty drives, e.g. V: + U: • Recommended to have at least two owners per shared folder	• Is appropriate for research dataset • **Is suitable for 'sensitive' (critical) data[4] where the data are used within Monash** • Suitable for small to large datasets • Backups occur 11am and 5pm (twice daily snapshots); previous versions are self-recoverable from last backup snapshot and are available for 30 days from when the file was last modified or deleted • External accounts are available for sharing non-sensitive data	• No direct allocations to undergraduates, but S: drive account owners (Monash staff) may grant access to students and other staff members without eSolutions Service Desk (SDO) mediation • No direct allocations to non-Monash users, but a request for an external account can be made through eSolutions Service Desk (SDO) (**for non-sensitive[4] data only**)
Local Drives • For Monash staff and students on the Monash Windows SOE o Desktop o My Documents, o My Pictures, not music or videos	• Windows 7 SOE: security is centrally managed – encryption uses BitLocker and TPM chipset • Backups: data saved to the Desktop, My Documents, My Pictures are synchronised with the network server in real time when online, immediate on next logon • Suitable for small to large datasets[1]	• **Not recommended for 'sensitive' (critical) data[4]** • Short-term storage for working data; not a permanent storage solution • Local hard drives (HDD) can fail from time to time
Monash 's Google Apps • For everyone with a Google account, or invited by University staff o Monash's Gmail o Google Drive o Google Sites o YouTube o Docs o Sheets o Slides	• Cloud-based: provides access to files locally and remotely for sharing and collaboration • Backups: three copies of virtual files are kept on the GFS (Google File System) across multiple Google data centres • Google Docs, Sheets and Slides are kept forever (all older versions are self-recoverable • Versions can be managed automatically • No cap on the number of versions kept (limited only by storage quota) • Google provides robust storage	• **Not suitable for 'sensitive' (critical) data[4]** • Data requiring long-term retention should be stored on other University storage services, such as shared drives • Suitable for small to medium datasets[1] – allocation is capped at 30GB and shared across the Apps i.e. cannot purchase additional space • Backups: process is proprietary, therefore undisclosed • Once a file is deleted and removed from the recycle/trash bin, it is gone forever
LaRDS **Mediated storage solution – contact the eSolutions Service Desk or the Monash eResearch Centre: merc@monash.edu** • For Monash staff and post graduate students • Access to other services through eResearch include: o High Performance Computing (HPC) and Visualisation - Monash Sun Grid (MSG), MASSIVE, NeCTAR Research Cloud, CAVE 2 o Collaboration tools – JIRA, Confluence, Sakai o Research Data Management – MyTardis, Research Data Storage Infrastructure (RDSI)	• Backups: 2-4 copies are kept across two data centres; 30 day history • For archiving modest to very large research datasets[1] • For supporting research data applications, instrumentation backup and visualisation • For the generation and analysis of research data	• 24-48 hour snapshot of changes • Mainly for archival data – may experience delays when restoring from archived files or datasets
Customised Solution For datasets that have complex information security requirements, particularly where data is shared outside Monash or has non-standard remote access requirements. • Allocations are provided case-by-case.	• Suitable for 'sensitive' (critical) data[4] • Suitable for small to large datasets[1]	• **Mediated storage solution for very specific data storage requirements. Service design (where applicable) is determined on a case-by-case basis; contact your eSolutions Service Desk (SDO) for an appropriate referral**
Transportable devices • Flash drives • Thumb drives • CDs • DVDs • USBs • Ext HDsww	• Readily available and cheap to buy	• Responsibility lies with the owner/purchaser, e.g. security and backups • Longevity is questionable; devices are prone to failure, theft and obsolescence, therefore not suitable for long-term storage • Not recommended for master copies of datasets • Encryption is required for 'sensitive' data[4] o BitLocker to Go (Win7) can be used on USBs and external hard drives. o Win XP and Mac users should consider encryption tools such as TrueCrypt (refer to tutorial) • Not suitable for large to very large datasets[1]
AARNet's CloudStor service Authcate access or invite by authcate user	• Up to 100 email recipients to share a file • Maximum file size is 100GB	• Not a storage solution – transmission of data only • **Not suitable for 'sensitive' (critical) data[4] as encryption only occurs during transmission** • Files and vouchers expire at a maximum of 20 days • Maximum 1 file per upload • Not suitable for master copies of data

Note: The research records management system "TRIM" has not been included in the storage matrix. TRIM is for Monash administrative staff only. For more details see the trim website: http://wwww.adm.monash.edu.au/records-archives/trim

1. Dataset size (guide only): 'small', up to 30GB; 'medium', between 30GB and 100GB; 'large', greater than 100GB up to 500GB; and, 'very large', greater than 500GB.
2. For the purposes of this document, active or working data are defined as data that require ongoing access for modification, analysis, compilation, etc. Archival storage solutions are more appropriate for 'end state' data.
3. Monash recommends use of the Monash Windows Standard Operating Environment (SOE) and can only provide limited support for data stored on non-SOE and non-Monash controlled environments. Mac users - the Mac OSX SOE is not yet certified.
4. For "sensitive" data classifications, refer to the following legislative definitions: Information Privacy Act 2000 (Vic), Privacy Act 1988; and, the Monash interpretation in the Electronic Information Security: Responsibilities, Classifications and Standards Procedures.

MONASH University Library

Digital data: personal storage

Memory sticks or portable hard disk drives (HDD)

- choose quality products from reputable manufacturers
- follow instructions for handling and storage
- make two copies, use ‘verify’ facilities and do a manual check for readability
- store a copy on a network drive
- ensure that private or confidential data is password-protected and/or encrypted

CDs and DVDs

- choose quality products from reputable manufacturers
- follow instructions for handling and storage
- burn all files at once, at low speed, and don’t use the computer while you are burning
- make two copies, use ‘verify’ facilities and do a manual check for readability
- store a copy on a network drive
- ensure that private or confidential data is password-protected and/or encrypted

Digital data: cloud storage

Choose quality products from reputable suppliers

read and understand their terms of use

- can they share information with third parties without consent?
- who owns the data stored in their system?
- who can access the data stored in their system?
- can they block access without consent?
- can they modify, edit or remove content without consent?
- what privacy laws apply?

- Follow 3-2-1 strategy, use ‘verify’ facilities and do a manual check for readability
- Ensure that private or confidential data is password-protected and/or encrypted

Print records and physical research objects

'Best practice' will differ depending on your data and the discipline

As a minimum, you should

- make sure that storage facilities are secure (e.g. lockable office or filing cabinet)
- be aware of environmental conditions and handling procedures that are most appropriate for your data
- develop strategies for keeping copies of original records and physical objects (if this is possible)

Controlled vocabularies (very useful)

- Lists of words or phrases used to classify (or tag)
- Subject headings, thesauri, taxonomies, ontologies, coding schemes
- Can be generic or very discipline-specific
- Wherever possible, use an existing standard, even if you need to adapt or customise it
- Classifying/tagging consistently will help you find and make sense of your data in future

More information from the library

Higher Degrees by Research Library Guide	http://guides.lib.monash.edu/hdr
Managing research data webpages	http://monash.edu/library/researchdata/
Guidelines	http://monash.edu/library/researchdata/guidelines/
Policy and governance	http://monash.edu/library/researchdata/about/policy
Copyright ownership and IP	http://monash.edu/library/researchdata/guidelines/ownership/
Data re-use	http://monash.edu/library/researchdata/guidelines/re-use/
Ethics	http://monash.edu/library/researchdata/guidelines/ethics/
Formats	http://monash.edu/library/researchdata/guidelines/formats/
Sharing	http://monash.edu/library/researchdata/guidelines/sharing/
Storage	http://monash.edu/library/researchdata/guidelines/storage/
Organising data	http://monash.edu/library/researchdata/guidelines/organising-data/
Deposit and publication	http://monash.edu/library/researchdata/guidelines/sharing/
Retention	http://monash.edu/library/researchdata/guidelines/retention/

Contacts

Data Management Coordinator

researchdata@lib.monash.edu.au

Your contact librarian

http://www.lib.monash.edu.au/contacts/faculty/

Monash e-Research Centre

merc@adm.monash.edu.au

Take away messages

- ☐ Understand the characteristics of research data you will use and create and how managing your research will benefit your research career
- ☐ Identify current relevant best practice in managing data and incorporate this in your work/future career
- ☐ Develop a research data management plan with a realistic timetable
  - Review regularly and revise when necessary
  - Avoid lengthy delays and distractions
- ☐ Employ the most appropriate resources, techniques and expertise to help you manage your data and your research