Introduction to research data planning for Higher Degree by Research students

This work is licensed under a Creative Commons Attribution-Noncommercial 2.5 Australia License.
Overview

- **Context**
  - Characterising research data
  - Managing data effectively
  - Where to get more information, advice and technical support

- **Research data management**
  - Research data planning checklist
  - Some legal, ethical, technical considerations
Characterising research data (1)

- Original material generated by research and/or unpublished resources analysed for research in original ways
- Very valuable part of your research
  - Validates your results
  - Enables others to build on your findings
Characterising research data (2)

- New (collected or created by you) and/or existing (sourced from somewhere else)
- Qualitative and/or quantitative
- Multiple formats
Common types of research data

- Statistics and measurements
- Results of experiments or simulations
- Observations e.g. fieldwork
- Survey results – print or online
- Interview recordings and transcripts, and coding applied to these
- Images, from cameras and scientific equipment
- Textual source materials and annotations
Discussion

- What kind of research are you doing?
- What types of data do you think you might generate?
Managing data well is part of being a responsible researcher

- Australian Code for the Responsible Conduct of Research (2007), Section 2

- Monash University Research Data Management Policy and Procedures (including HDR procedures)
Effective management of data

Ownership & rights

National & institutional policy frameworks

Copyright & IP in new data

Using existing data

Ethics

File formats

Storage & backup

Sharing & controlling access

Organising & documenting data

Publishing & dissemination

Long-term retention

Secure destruction

Leaving Monash
Data planning checklist

- Helps you develop a plan for how you will manage your data
- Will be of most benefit completed early, but is a work in progress
Complete quickly and easily using multi-choice boxes.

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

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

- Pick up a copy from your campus Library
- Or download from the web: www.researchdata.monash.edu.au/resources/dataleaflet.pdf
Research data management

Monash University is committed to improving the way research data in all formats is created, stored, managed and disseminated. This website provides guidance on common research data management issues and highlights the portfolio of advisory and technical support services that are available to researchers.

Guidelines
- Data management planning
- Ownership, copyright and IP
- Ethical requirements
- Durable formats
- Storage and backup
- Sharing data and controlling access
- Documentation and metadata
- Retention and disposal
- Secure destruction
- Deposit in a repository or archive

Latest
- exPERT seminars for HDR students
  1. Introduction to Data Management
  Registrations opening soon for Clayton and Caulfield campus sessions in May-June 2010
  Registrations opening soon for Clayton and Caulfield campus sessions in May-June 2010
- Seeking a customised session or individual consultation? Contact the University’s Data Management Coordinator
Other related exPERT seminars

- Introduction to Intellectual Property and Copyright
- Ethics (various modules)
- Software packages for statistics
Advice - Library

- Data Management Coordinator
  researchdata@lib.monash.edu.au

- Faculty contact librarians
  http://www.lib.monash.edu.au/contacts/faculty/
Monash e-Research Centre (MeRC) complements faculty-based IT support, with a focus on supporting researchers:

- Storage and backup
- Organising and documenting data
- Collaboration
A. Copyright, ownership and IP
Copyright basics (1)

- 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
- Does not need to be registered in Australia – it applies automatically
Copyright basics (2)

- If your research is conducted in Australia, then Australian copyright law will apply.
- In Australia copyright will apply to data in most cases - 'originality' only requires that some skill, ingenuity and labour has been involved, not that the thought or idea is especially innovative.
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:
  www.mrgs.monash.edu.au/research/staff/ip/assignment.doc
Situations in which students assign the IP

- The university makes a specific contribution AND you make or contribute to a patent-worthy discovery or invention
- You are provided with background IP owned by the university or your supervisor
- Funding agreement between the university and a third party states that the outcomes of the research are owned by someone other than you
Using third party data (1)

- 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
Using third party data (2)

- ‘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 written agreements or permissions (e.g. emails, letters from the copyright holders)
Activity

- Explore the terms and conditions for re-use of data from the following sources:
  - Australian Bureau of Statistics census data
  - Victoria Water Resource Data Data Warehouse
  - Bureau of Meteorology Climate Data Online
  - Worldwide Protein Databank
  - State Library of Victoria Picture Collection
  - OECD iLibrary statistics
Ownership, copyright and IP: where to go for more information

- **Copyright at Monash website**
  http://www.copyright.monash.edu.au/

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

  http://www.oaklaw.qut.edu.au/reports
B. Ethical requirements
Ethical requirements: privacy, confidentiality, cultural sensitivity

- Particularly important for research involving humans e.g. surveys, interviews, analyses of personal data, medical research
- Addressed through the human ethics application process - your supervisor will guide you through
- Regular MRGS exPERT seminars on ethics
Privacy

- Monash Privacy Policy and other privacy legislation:
  - the Information Privacy Act 2000 (Vic)
  - the Health Records Act 2001 (Vic)
  - the Privacy Amendment (Private Sector) Act 2000

- Privacy Coordinator for your faculty or unit

- University Privacy Officer
  privacyofficer@adm.monash.edu.au
Confidentiality

- *National Statement on Ethical Conduct in Research Involving Humans* – esp. Chapter 3.2 on Databanks

- Consent processes must include details of
  - how data will be stored
  - whether subjects will be identifiable
  - purposes for which the data will be used
Cultural sensitivity

- Research involving indigenous people may have special requirements
  - decisionmaking and communicating about methods of collecting data
  - where data is stored and how it is accessed
  - Intellectual and cultural property rights

- **AIATSIS Guidelines for Ethical Research in Indigenous Studies** are a good starting point:
  
Ethics and data management

- Potential tensions between ethical requirements and other requirements for data sharing
- Important to consider these issues early
- Researchers who want to share or re-use data in future can consider strategies like:
  - Seek informed consent for proposed data sharing or re-use
  - Protect privacy through anonymising data
Methods for anonymising data: quantitative

- Identity can be disclosed directly or indirectly
- Remove direct identifiers, e.g. name
- Broaden the meaning of variables, e.g. age groups instead of date of birth, broad professional category rather than job title
- Relational data and geospatial data may be identifying too so take extra care
Methods for anonymising data: qualitative

- Anonymise at transcription / write-up
- Use pseudonyms or replacements
- Keep a log of substitutions - store separately from files
- May need to keep non-anonymised versions too
- Audio / video manipulation is possible but difficult and expensive – better to obtain consent up front if you want to share this kind of material
More information about consent, confidentiality and anonymising data

C. Durable formats
Durable digital 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 by a standards agency e.g. 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.
D. Storage and backup
Digital data: university networks (1)

- Storing your research data on university networks 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
Digital data: university networks (2)

Options include:

1. Monash Large Research Data Store (LaRDS) – run by central ITS / Monash e-Research Centre
2. Faculty- or school-based IT services – provision will depend on the faculty or unit
LaRDS (1)

http://www.monash.edu.au/eresearch/services/lards

- LaRDS is Monash’s central research data store
- Use of LaRDS is welcomed and encouraged by all Monash researchers (including HDR students)
- No cost for standard use
LaRDS (2)

- NOT just a single…service, ‘file system’, interface or software application
- LaRDS is a common infrastructure supporting many different applications and access modes
- The e-Research Centre can help you work out which way of accessing LaRDS is best for you
How to get access to LaRDS

- Email merc@adm.monash.edu.au – include as much of the following information as you can
  - Brief research description (~150 words)
  - Mac / Windows PC / Linux
  - Data size, data type and data security requirements
  - Duration of your project
  - Users - just you / you and your supervisor / group? Anyone outside Monash?
  - How often you will access the data, and from where – office, lab, home, in the field?
  - How you heard about LaRDS (e.g. exPERT seminar)
Digital data: personal storage (1)

- CDs, DVDs, USB sticks, portable hard drives - widely used, cheap and portable but not recommended for master copies
  - may not be large enough for all your data
  - failure and errors - bad ‘burning’ / writing to disk, physical damage (shocks, magnetism), sensitivity to environmental conditions
  - security risk - regularly misplaced or lost, data not always password-protected or encrypted
Digital data: personal storage (2)

- If using CDs and DVDs, learn about best practice: e.g. [http://www.jiscdigitalmedia.ac.uk/](http://www.jiscdigitalmedia.ac.uk/)
  - choose quality products from reputable suppliers
  - 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
  - ensure that private or confidential data is password-protected and/or encrypted
Backing up digital data

- LaRDS is automatically backed up to a tape library nightly in two physical locations
- Backup on faculty drives may vary - check frequency, number of copies, and locations
- Online remote backup services – check terms and conditions carefully
Print and physical data

- ‘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
  - keep additional copies (if this is possible)
E. Sharing & controlling access
Controlling access

- Digital, e.g.
  - Use systems protected with username/password as a minimum

- Print and physical, e.g.
  - lock rooms, filing cabinets and safes
  - secure laptops and drives
Enabling access

- Sharing with supervisors, other students, and more widely – *where appropriate*
- Monash e-Research Centre tools are available that can help with this – Google apps (Monash version) is also an option
- Many free tools are available but may not guarantee security of your data
F. Documentation and metadata
Documentation and metadata

- How will you (or others) need to find and make sense of your data in future?
- Are there standards in your discipline for documentation and metadata (structured information about the data)?
- How can you make your data as ‘self-documenting’ as possible? (e.g. meaningful filenames, titles, column headings)
Documentation

- Provides context for the data
- Could include information such as:
  - why data was collected or created
  - how data was collected, e.g. instruments used
  - structure, e.g. labels and descriptions for elements, coding or classification schemes
  - quality control measures
  - any other information that would help others analyse and interpret the data e.g. user guides or manuals
Metadata

- A subset of documentation
- Structured information about the data - usually conforming to a standard or schema
- Descriptive metadata - identifies the data and enables it to be searched and browsed
- Administrative and technical metadata - enables the data to be better managed by capturing information such as creation dates, file formats and access restrictions
Controlled vocabularies

- Lists of words or phrases used to classify (or tag) your resources
- May be called subject headings, thesauri, taxonomies, ontologies, coding schemes
- Can be generic or very discipline-specific
- Wherever possible, use an existing standard
- Classifying/tagging consistently will help you find and make sense of your data in future
Example: Social Science

- Many data archives use the Data Documentation Initiative (DDI) standard
- About DDI - http://www.ddialliance.org/
Example: Humanities

- Different standards – find out what is relevant to your discipline
- Dublin Core is one widely used standard
Example: Science

- Different standards – find out what is relevant to your discipline
- Content Standard for Digital Geospatial Metadata is widely used - [www.fgdc.gov/metadata/geospatial-metadata-standards](http://www.fgdc.gov/metadata/geospatial-metadata-standards)
- E.g. *Commercial and Research Fish Catch database held by the Australian Antarctic Division* - [http://tinyurl.com/2996dk5](http://tinyurl.com/2996dk5)
G. Retention periods
Minimum retention periods

- Data must be kept for a *minimum* of 5 years from the end of the project or publication of the results.
- In some cases, retention periods are longer:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medical research involving clinical trials</strong></td>
<td>15 years</td>
</tr>
<tr>
<td><strong>Psychological testing or intervention (adults)</strong></td>
<td>7 years</td>
</tr>
<tr>
<td><strong>Psychological testing or intervention (children)</strong></td>
<td>25 years after date of birth</td>
</tr>
</tbody>
</table>
Permanent retention

- Data of archival interest may be kept permanently
- Characteristics of archival data may include:
  - Controversial
  - Innovative technique used for the first time
  - Research of known public interest
  - Observational data that would be costly or impossible to reproduce
  - Data of high interest to researchers in same or other disciplines
SCIENTISTS at the University of East Anglia (UEA) have admitted throwing away much of the raw temperature data on which their predictions of global warming are based.

It means that other academics are not able to check basic calculations said to show a long-term rise in temperature over the past 150 years.

The UEA’s Climatic Research Unit (CRU) was forced to reveal the loss following requests for the data under Freedom of Information legislation.

The data were gathered from weather stations around the world and then adjusted to take account of variables in the way they were collected. The revised figures were kept, but the originals — stored on paper and magnetic tape — were dumped to save space when the CRU moved to a new building.

The admission follows the leaking of a thousand private emails sent and received by Professor Phil Jones, the CRU’s director. In them he discusses thwarting climate sceptics seeking access to such data.

In a statement on its website, the CRU said: “We do not hold the original raw data but only the value-added (quality controlled and homogenised) data.”
H. Publishing & disseminating
Making data available

- Consider the future use and value of your data – to you, other researchers and wider audiences
- The Code suggests making your data widely available - *if there are no restrictions*
- Depends to some extent on who owns the data
- Decisions made early in your project (e.g. around confidentiality) may restrict your ability to make your data available later
Publishing & research impact

- International trend towards more open access to data in archives & repositories
- Some journals now expect supporting data to be available for peer review or on publication
- Some early evidence that when data is publicly available, the publications get cited more
- Increasing use of multimedia as part of publications and presentations
- Reaching out to audiences that do not have access to expensive journals
Repositories and archives

- Will differ depending on discipline and data types but are likely to have:
  - preferred data formats
  - minimum standards for documentation and metadata
  - agreements or licences that ensure that the data meets legal and ethical requirements
- Find out if there are archives or repositories for data from your discipline, and be sure that you understand deposit requirements
Some Australian examples

- Australian Social Science Data Archive (ASSDA)
- Aboriginal Studies Electronic Data Archive (ASEDA)
- BlueNet - the Australian Marine Science Data Network
- Pacific and Regional Archive for Digital Sources in Endangered Cultures [linguistics]
Some international examples

- Biomedical Informatics Research Network Data Repository
- EarthChem
- GenBank
- Inter-University Consortium for Political and Social Research
- International Spectroscopic Data Bank
- Knowledge Network for Biocomplexity
- Petrological Database of the Ocean Floor
- System for Earth Sample Registration
- TreeBASE: A Database of Phylogenetic Knowledge
- World Data Center for Geophysics & Marine Geology
- World Wide Protein Data Bank (PDB)
Making other long-term arrangements

- Even if data cannot go in a repository or archive, you need to make long-term arrangements.
- Depends to some extent on who owns the data.
- Discuss options with your supervisor:
  - Faculty / school / supervisor keeps a copy
  - Keep data yourself – follow the LOCKSS principle (Lots of Copies Keeps Stuff Safe!)
  - Transfer data to another institution
Secure destruction

- May be required on ethical grounds or because data is no longer of value
- Must be carried out using an irreversible method
- Extra care must be taken when dealing with records that contain sensitive information
What to do with your planning checklist

- 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
Things to remember

- Your research data is a precious asset
- You need to keep data \textit{at least} 5 years after the end of your project
- Monash has many services and people that can help you – but it’s up to you to make the most of them
- Find out more through further exPERT seminars, web resources and getting in touch