Managing Digital Research Data at ANU

This work by The Australian National University is licensed under a Creative Commons Attribution 4.0 International license
Presenter: Zixiao Wang
zixiao.wang@anu.edu.au
repository.admin@anu.edu.au
Acknowledgment of Country:

We acknowledge and celebrate the Ngunnawal people on whose traditional lands we meet today, and pay our respect to elders, past, present and emerging.
ANU Open Research Repository

- Meets open access requirements for ARC/NHMRC grants
- Mandated location for theses
ANU Research Data Management

Contents:

• Definitions of Data and Data Management
• Data Management Life Cycle
• Benefits & Requirements
• Methods of Data Management
• ANU's Data Management Services
• Writing a Data Management Plan
Digital Research Data

Any data that is created during research that can be stored on a computer.

This includes field notes and non-digital images as they can be converted to digital images.

It is not:

- physical data such as biological specimens, soil samples etc.
- Admin data, research publications, course resources etc.
# Data Management

## Organisation
- Bibliography Management
- File transfers & remote access
- Synchronisation
- Collaboration
- Version Control

## Administration
- Backups
- Validation & Authentication
- Documentation
- Access Control
- Security

## Archiving
- Sharing Methods
- Licensing
- Formats & Standards
- Access Restrictions
- Metadata
- Archiving
- FAIR Principles
Data Life Cycle

DataOne Primer
Why data management?

- Work more efficiently with data
- Protect data against loss or improper access
- Greater exposure, enhanced reputation and prestige
- Reproducibility
Why data management?

• Comply with ethical and privacy regulations
• Funding agencies
  – Data relating to publications must be kept for at least 5 years
• Journals require sharing of data, e.g. Nature, Science, PLoS, etc.
Methods of Data Management

• Data organisation
  – Bibliography management: EndNote, RefWorks, Mendeley, Zotero
  – Version control & file naming: TortoiseSVN, GitHub, BitBucket
  – File transfers: CloudStor, OneDrive, SharePoint
  – File storage: CloudStor, OneDrive, OneNote, HomeDrive, ANU eNotebooks, SharePoint, NCI

• Data administration
  – Backups: frequency etc
  – access control

• Data archiving and sharing
  – creative or science commons licenses
ANU's Data Management Services

- Cybersafety at home
- ANU Virtual Information Commons
- Local IT Support Staff (LITSS)
- Information Technology Services (ITS)
  - HomeDrive, OneDrive, SharePoint
- ANU Library (Digital Literacy Training, Open Research)
- Statistical Consulting Unit
- Archives: Data Commons, Open Research, ADA
- Supercomputer (NCI)
ANU Data Storage

- CloudStor
- SharePoint
- OneDrive
- OneNote
- ANU eNotebooks
- National Computation Infrastructure (NCI)
ANU Data Archives

- Open Research – ANU's institutional repository. Good for publications, theses, images

- ADA – Australian Data Archive. Social-science, medical, and economics datasets. Analyse datasets online with the NESSTAR webapp

- Data Commons – ANU data archiving service
ANU Data Commons

- Users can deposit data for long term storage
- Can mint DOIs for datasets
- Users can self archive via web interface
- Data can be restricted or open
- Metadata can be restricted or open
ANU Data Commons (cont.)

• As long as storage is accessible via URL, Data Commons can link to it – create a reference
• Data for work in progress can be stored
• Large datasets (e.g. over 10 GB) can be uploaded via command line
HomeDrive

- 4.5GB Storage available
- Access via web browser (ANU login): myfiles.anu.edu.au
Office 365/SharePoint

- A web-based collaborative platform, available on MS Office 365, used to co-author documents; Manage versions; Apply workflows; Discover and share information
- OneDrive stores your personal documents
Digital Literacy Training & Support

• Introduction to Library research
• Advanced Research: using ANU Library Databases
• EndNote
• Word for academic writing workshops
• Research data management
• NVivo

Book online at ANU Library: anulib.anu.edu.au/training-register

• Digital Essentials
• SPSS
• Publishing with LaTeX
• Working with data in Excel

Online recordings and support: anulib.anu.edu.au/research-learn
IT Security @ ANU

• Intercept X for ANU managed machines
• Anti-Virus comparison, ANU users are encouraged to compare and install on personal devices.

services.anu.edu.au/information-technology/software-systems/sophos-anti-virus
The National Computational Infrastructure

• Fastest Supercomputer in Australia
• All staff/students can apply for access
• For complex computational or large data projects
• Dataset hosting for every large datasets of national significance
Benefits of Digital Archiving

- Long term storage of end-of-life data
- Converted to formats that are durable and free
- Owner sets Access Restrictions
  - Unrestricted: anyone can download
  - Registered: requires name and affiliation
  - Requested: users submit a request for data
  - Closed: no access
Let’s avoid this…

Data Sharing and Management Snafu in 3 Short Acts

By Karen Hansen, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries

www.youtube.com/watch?v=66oNv_DJuPc
DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. Find out more.

1. REGISTER
   Get your unique ORCID identifier Register now!
   Registration takes 30 seconds.

2. ADD YOUR INFO
   Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).

3. USE YOUR ORCID ID
   Include your ORCID identifier on your Webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

Additional information: Your ORCID iD and ANU anulib.anu.edu.au/howto
Writing a Data Management Plan

The Australian National University is a participating institution of the DMPTool service. ANU recommends the use of the default DMPTool template.

To complete a plan, you will need to consider the following:

**Data Collection**
- What data will you collect or create?
- How will the data be collected or created?

**Documentation and Metadata**
- What documentation and metadata will accompany the data?

**Ethics and Legal Compliance**
- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights issues?

**Storage and Backup**
- How will the data be stored and backed up during the research?
- 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?

**Data Sharing**
- How will you share the data?
- Are any restrictions on data sharing required?

**Responsibilities and Resources**
- Who will be responsible for data management?
- What resources will you require to deliver your plan?

*provided by the California Curation Centre of the California Digital Library*
Data Management Plan

- From 2020 ARC will require that data management plans are in place prior to the commencement of the project.
  
  arc.gov.au/research-data-management

- ANU has developed an online DMP platform:
  dmptool.org
Metadata is "data that provides information about other data"

- Descriptive
  - What is it about?
  - Title, time, author, keywords
  - Relations to other data objects
- Administrative (Access & Rights)
  - Ownership and use permissions
- Provenance
  - Where does it come from?
  - History of changes to the data, versions
- Standards/Schemas
Metadata to Consider:
Who, What, Where, When, Why, How

• Name of the data set and data files
• Date of creation and last modification
• Software used to create file (including version)
• Data processing performed
• Who collected the data
• Contact information of responsible party
• Sponsor or funding agencies
• Why the data were collected (abstract; keywords; controlled vocabulary); when and where
• Instrumentation; experimental conditions; calibrations
• Units of measure
• Taxonomic details
• Known problems that limit data use
• How to cite the data set
Example of Metadata

- Records from Research Data Australia:
  - researchdata.ands.org.au/time-series-plankton-loganapos-dam/445000
- A record from OpenResearch@ANU
  - openresearch-repository.anu.edu.au/handle/1885/138114?mode=full
- The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) paradisec.org.au/info.html
Further Information

ANU Research Data Management Manual
ql.anu.edu.au/cco9

LibGuide Data Management
ql.anu.edu.au/data
Resources

LibGuide Data Management: ql.anu.edu.au/data
ARDC: ands.org.au
RE3Data.org: re3data.org
ANU Research data management resources: anulib.anu.edu.au/research-learn/research-data-data-management
ANU Talk data to me: anulib.anu.edu.au/publishing
Data Repositories: journals.plos.org/plosone/s/data-availability
Google Dataset Search (Beta): toolbox.google.com/datasetsearch
Top Tips

✓ Have a Data Management Plan
✓ Back up your data
✓ Register an ORCID
MOST IMPORTANT

BACK UP YOUR DATA

You can’t create a data management plan, share or reproduce your results if you have lost your data!